Tracing Genetic Exchange and Biogeography of *Cryptococcus neoformans* var. *grubii* at the Global Population Level

Johanna Rhodes,*,[1] Christopher A. Desjardins,[†,1] Sean M. Sykes,[†] Mathew A. Beale,*,[‡,§] Mathieu Vanhove,* Sharadha Sakthikumar,[†] Yuan Chen, Sharvari Gujja,[†] Sakina Saif,[†] Anuradha Chowdhary,[††] Daniel John Lawson,[‡‡] Vinicius Ponzio,[§§] Arnaldo Lopes Colombo,[§§] Wieland Meyer,***,[†††] David M. Engelthaler,[‡‡‡] Ferry Hagen,[§§§,****] Maria Teresa Illnait-Zaragozi,[††††] Alexandre Alanio,[‡‡‡‡] Jo-Marie Vreulink,[§§§§] Joseph Heitman,***** John R. Perfect,** Anastasia P. Litvintseva,***** Tihana Bicanic,[‡] Thomas S. Harrison,[‡] Matthew C. Fisher,*,[2] and Christina A. Cuomo[†,2]**

*Department of Infectious Disease Epidemiology, Imperial College London, W2 1PG, United Kingdom, [†]Infectious Disease and Microbiome Program, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, [‡]Institute of Infection and Immunity, St. George's University London, WC1E 6BT, United Kingdom, [§]Infection Genomics, Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom, **Division of Infectious Diseases, Department of Medicine, and *****Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, [††]Department of Medical Mycology, Vallabhbhai Patel Chest Institute, University of Delhi, 110007, India, [‡‡]Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, BS8 1TH, United Kingdom, [§§]Division of Infectious Diseases, Federal University of São Paulo, 04039-032, Brazil, ***Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Sydney Medical School–Westmead Hospital, Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Westmead Institute for Medical Research, 2145, Australia, [†††]Mycology Laboratory, Evandro Chagas National Institute of Infectious Diseases, Oswaldo Cruz Foundation, Rio de Janeiro, 21040-360, Brazil, [‡‡‡]TGen North, Translational Genomics Research Institute, Flagstaff, Arizona 86005, [§§§]Department of Medical Microbiology and Infectious Diseases, Canisius-Wilhelmina Hospital, 6532SZ Nijmegen, The Netherlands, ****Centre of Expertise in Mycology, Radboudumc/Canisius-Wilhelmina Hospital, 6532SZ Nijmegen, The Netherlands, [††††]Departamento Bacteriología-Micologia, Centro de Investigación, Diagnóstico y Referencia, Instituto de Medicina Tropical Pedro Kourí, La Habana, 601, Cuba, [‡‡‡‡]Laboratoire de Parasitologie-Mycologie, Assistance Publique–Hôpitaux de Paris, Groupe Hospitalier Saint-Louis-Lariboisière-Fernand-Widal Paris, 75010, France; Université Paris Diderot, Sorbonne Paris Cité, 75010, Paris, France; Unité de Mycologie Moléculaire, Institut Pasteur, Centre National de la Recherche Scientifique, Centre National de Référence Mycoses Invasives et Antifongiques, URA3012, 75015, Paris, France, and [§§§§]Department of Microbiology, Stellenbosch University, 7600, South Africa

ORCID IDs: 0000-0002-1338-7860 (J.R.); 0000-0002-3160-4734 (C.A.D.); 0000-0002-4740-3187 (M.A.B.); 0000-0002-2028-7462 (A.C.); 0000-0002-5311-6213 (D.J.L.); 0000-0001-9933-8340 (W.M.); 0000-0002-5622-1916 (F.H.); 0000-0001-9726-3082 (A.A.); 0000-0001-5446-1396 (J.-M.V.); 0000-0001-6369-5995 (J.H.); 0000-0003-0019-8638 (J.R.P.); 0000-0002-1862-6402 (M.C.F.); 0000-0002-5778-960X (C.A.C.)

ABSTRACT *Cryptococcus neoformans* var. *grubii* is the causative agent of cryptococcal meningitis, a significant source of mortality in immunocompromised individuals, typically human immunodeficiency virus/AIDS patients from developing countries. Despite the worldwide emergence of this ubiquitous infection, little is known about the global molecular epidemiology of this fungal pathogen. Here we sequence the genomes of 188 diverse isolates and characterize the major subdivisions, their relative diversity, and the level of genetic exchange between them. While most isolates of *C. neoformans* var. *grubii* belong to one of three major lineages (VNI, VNII, and VNB), some haploid isolates show hybrid ancestry including some that appear to have recently interbred, based on the detection of large blocks of each ancestry across each chromosome. Many isolates display evidence of aneuploidy, which was detected for all chromosomes. In diploid isolates of *C. neoformans* var. *grubii* (serotype AA) and of hybrids with *C. neoformans* var. *neoformans* (serotype AD) such aneuploidies have resulted in loss of heterozygosity, where a chromosomal region is represented by the genotype of only one parental isolate. Phylogenetic and population genomic analyses of isolates from Brazil reveal that the previously "African" VNB lineage occurs naturally in the South American environment. This suggests migration of the VNB lineage between Africa and South America prior to its diversification, supported by finding ancestral recombination events between isolates from different lineages and regions. The results provide evidence of substantial population structure, with all lineages showing multi-continental distributions; demonstrating the highly dispersive nature of this pathogen.

KEYWORDS *Cryptococcus*; hybridization; phylogeography; recombination; selection; genome sequence

THE environmental basidiomycetous yeast *Cryptococcus neoformans* is capable of causing invasive fungal infections primarily in immunocompromised individuals. Meningitis is the most serious manifestation of cryptococcosis. The human immunodeficiency virus (HIV)/AIDS pandemic increased the population of these susceptible individuals and led to an increase in *C. neoformans* infection rates (Day 2004). *C. neoformans* is the leading cause of mortality in HIV/AIDS patients worldwide, particularly in sub-Saharan Africa, where approximately half a million deaths occur annually (Park *et al.* 2009). While cryptococcal infection rates in HIV-positive individuals have declined due to highly active antiretroviral therapy (HAART), new estimates continue to suggest there are >100,000 deaths/year (Rajasingham *et al.* 2017). Recent data also suggest that the incidence of cryptococcosis has plateaued at a high number, despite HAART availability. Furthermore, the increasing number of people living with other immunodeficiencies, including transplant and cancer patients, represents a growing population at risk for cryptococcosis (Maziarz and Perfect 2016).

There are three major serotypes of *C. neoformans* distinguished by different capsular antigens, which include two separate varieties (*C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans*, serotypes A and D, respectively) and a hybrid between the two (serotype AD). While *C. neoformans* isolates are primarily haploid, diploid AD hybrid isolates consisting of both serotype A (*C. neoformans* var. *grubii*) and serotype D (*C. neoformans* var. *neoformans*) have been isolated from both clinical and environmental sources mostly in Europe (Franzot *et al.* 1999; Cogliati 2013; Desnos-Ollivier *et al.* 2015). Serotype A isolates are the most common cause of infection, accounting for 95% of all *C. neoformans* infections globally (Casadevall and Perfect 1998; Heitman *et al.* 2011). Genomes of serotype A and D isolates differ by 10–15% at the nucleotide level (Loftus *et al.* 2005; Kavanaugh *et al.* 2006; Janbon *et al.* 2014), and laboratory crosses of A and D isolates are possible but show reduced viability of meiotic spores (Lengeler *et al.* 2001; Vogan and Xu 2014).

C. neoformans var. *grubii* can be divided into three molecular types, or lineages: VNI, VNII, and VNB (Meyer *et al.* 1999, 2009; Litvintseva *et al.* 2006). The VNI and VNII lineages are isolated globally, while the VNB lineage is predominantly located in sub-Saharan Africa (Litvintseva *et al.* 2006), although there is some evidence for VNB occurring in South America (Bovers *et al.* 2008; Ngamskulrungroj *et al.* 2009) and in the United States, Italy, and China in AD hybrid isolates (Litvintseva *et al.* 2007). Apart from clinical isolation, the VNI lineage is primarily associated with avian excreta (Nielsen *et al.* 2007; Lugarini *et al.* 2008), while the VNB lineage is found mostly in association with specific tree species, predominantly mopane trees (Litvintseva *et al.* 2011; Litvintseva and Mitchell 2012). This and recent studies have shown that VNI infections are associated with urbanized populations where an avian-associated reservoir, pigeon guano, is also found; while the VNB lineage is widely recovered in the African arboreal environment (Litvintseva *et al.* 2011; Vanhove *et al.* 2017).

Mating in *C. neoformans* occurs between cells of opposite mating types (*MAT***a** and *MAT*α) (Kwon-Chung 1975, 1976), although unisexual mating can also occur (Lin *et al.* 2005). *MAT*α isolates are capable of unisexual mating both within and between the two serotypes (Lin *et al.* 2005, 2007), and recombination was shown to occur at similar levels in bisexual and unisexual mating in serotype D isolates (Sun *et al.* 2014; Desnos-Ollivier *et al.* 2015). Due to the rarity of *MAT***a** isolates of both serotypes in the environment (Lengeler *et al.* 2000a; Viviani *et al.* 2001; Litvintseva *et al.* 2003), unisexual mating may have evolved to enable meiotic recombination and genetic exchange between isolates. Several studies have found evidence of recombination within VNI, VNII, and VNB populations, although not between these lineages (Litvintseva *et al.* 2003, 2005; Bui *et al.* 2008).

An additional level of genome diversity detected in *C. neoformans* var. *grubii* includes the presence of cryptic diploid isolates and variation in the copy number of individual chromosomes or regions. Close to 8% of *C. neoformans* var. *grubii* global isolates appear diploid; these isolates contain the *MAT*α locus and many appear autodiploid, thought to result either from endoreduplication or self-mating (Lin *et al.* 2009). While the vast majority of serotype A or D isolates appear haploid, individual chromosomes can be present at diploid or triploid levels (Hu *et al.* 2011). For chromosome 1, a specific advantage of aneuploidy is copy number amplification of the azole drug targets or efflux transporters, associated with drug resistance (Sionov *et al.* 2010). While the specific selective advantage of other chromosomal aneuploidies is unknown, same-sex mating of *MAT*α isolates generates aneuploid progeny at high frequency, some of which also exhibit azole resistance (Ni *et al.* 2013). Titan cells, polyploid yeast cells produced in the lung of infected animals, also generate aneuploid progeny under stress conditions (Gerstein *et al.* 2015).

Previous studies examining the global population structure of *C. neoformans* var. *grubii* have used typing methods for a few genetic loci or focused on particular geographic regions or countries (de Oliveira *et al.* 2004; Litvintseva *et al.* 2006;

doi: https://doi.org/10.1534/genetics.117.203836
Manuscript received May 15, 2017; accepted for publication June 28, 2017; published Early Online July 5, 2017.
Available freely online through the author-supported open access option.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.203836/-/DC1.
[1]These authors contributed equally to this work.
[2]Corresponding authors: Department of Infectious Disease Epidemiology, Imperial College London, W2 1PG, United Kingdom. E-mail: matthew.fisher@imperial.ac.uk; and Broad Institute, 7 Cambridge Center, Cambridge, MA 02142. E-mail: cuomo@broadinstitute.org

Hiremath *et al.* 2008; Khayhan *et al.* 2013). Recent approaches have applied whole-genome sequencing (WGS) to trace the microevolution of *Cryptococcus*, identifying variation that occurs during the course of infection (Ormerod *et al.* 2013; Chen *et al.* 2017; Rhodes *et al.* 2017) or in the environment (Vanhove *et al.* 2017). Here, we use WGS of 188 isolates to provide a comprehensive view of the population variation between the three major lineages. The sequenced isolates were selected to represent the diversity of *C. neoformans* var. *grubii*, including each of the three major lineages and global geographic sampling. We identify contributions to genomic diversity generated through interlineage meiotic exchange to create haploid hybrids, generation of AD diploid hybrids, and regional copy number amplification. Furthermore, we finely analyze the phylogenetic relationships and trace the evolution of *C. neoformans* var. *grubii* at the global population level.

Materials and Methods

Isolate selection

A total of 188 *C. neoformans* var. *grubii* isolates were selected from previous studies, which include 146 clinical isolates, 36 environmental isolates, 4 animal isolates, and 2 isolates of unknown isolation source. These isolates were collected from 14 different countries: Argentina, Australia, Botswana, Brazil, China, Cuba, France, India, Japan, South Africa, Tanzania, Thailand, Uganda, and USA (Supplemental Material, Table S1). Most of the clinical isolates were isolated from the cerebrospinal fluid of patients. A total of 8 of the 36 environmental isolates were isolated from pigeon guano, and most of the remaining isolates were collected from mopane and other tree species.

Details of clinical trials and ethical review

French isolates were collected during the Crypto A/D study (Dromer *et al.* 2007). The study was approved by the local ethical committee and reported to the French Ministry of Health (registration no. DGS970089). For clinical trials undertaken in South Africa (Bicanic *et al.* 2007, 2008; Jarvis *et al.* 2012; Loyse *et al.* 2012) and Thailand (Brouwer *et al.* 2004), ethical approval was obtained from the Wandsworth Research Ethics Committee covering St. George's University of London. Local ethical approval was obtained from the University of Cape Town Research Ethics Committee in South Africa, and the ethical and scientific review subcommittee of the Thai Ministry of Public Health. Clinical isolates from India were collected during routine diagnostic service; local ethical approval was obtained from the Institutional Ethical Committee of Vallabhbhai Patel Chest Institute, University of Delhi, India.

Fluconazole sensitivity testing

Fluconazole minimum inhibitory concentrations (MICs) were determined for two isolates by the National Health Laboratory Service laboratory in Green Point, Cape Town, using the E-test method (Biomerieux) (Bicanic *et al.* 2006).

DNA isolation and sequencing

Each yeast isolate was recovered from a freezer stock and purely cultured on a yeast, peptone, dextrose (YPD) or Sabouraudthetic dextrose (SD) agar plate for 48–60 hr. Next, a single colony was inoculated to another YPD plate and cultured for 24 hr. Approximately 100 μl of yeast cells were used for DNA isolation using the MasterPure Yeast DNA Purification Kit (Epicenter, Madison, WI) according to the manufacturer's instructions. Alternatively, a single colony was inoculated into 6 ml YPD broth supplemented with 0.5 M NaCl and cultured for 40 hr at 37°, prior to extraction using the MasterPure Yeast DNA Purification Kit (Epicentre) as previously described (Rhodes *et al.* 2017).

DNA was sequenced using Illumina technology. For each isolate, a small insert library was constructed and used to generate between 14 and 150 million 101-bp, paired-end reads per isolate, which resulted in 56- to 603-fold average coverage of reads aligned to the H99 genome. In addition, large insert libraries were constructed for 15 isolates (Table S4) and also used to generate 101-bp, paired-end reads. Isolates were sequenced at Imperial College London and the Broad Institute (Table S1).

Read alignment, variant detection, and ploidy analysis

Illumina reads were aligned to the *C. neoformans* var. *grubii* reference genome H99 (Janbon *et al.* 2014) using the Burrows–Wheeler Aligner (BWA) 0.7.12 mem algorithm (Li 2013) with default parameters. BAM files were sorted and indexed using Samtools (Li *et al.* 2009) version 1.2. Picard version 1.72 was used to identify duplicate reads and assign correct read groups to BAM files. BAM files were locally realigned around indels using GATK (McKenna *et al.* 2010) version 3.4-46, "RealignerTargetCreator," and "IndelRealigner."

SNPs and indels were called from all alignments using GATK version 3.4-46, "HaplotypeCaller" in GVCF mode with ploidy = 1, and GenotypeGVCFs was used to predict variants in each isolate. All VCFs were then combined and sites were filtered using VariantFiltration with QD $<$ 2.0, FS $>$ 60.0, and MQ $<$ 40.0. Individual genotypes were then filtered if the minimum genotype quality was $<$50, percent alternate allele was $<$0.8, or depth was $<$10.

In examining isolates with a high proportion of sites that were removed by these filters, inspection of the allele balance supported that these isolates were diploid. For heterozygous diploid isolates, haplotypeCaller was run in diploid mode. VariantFiltration was the same, with the added filter of ReadPosRankSum $<$ −8.0. There was no allele depth filter for individual genotype filtration, but otherwise filtration was the same as for haploid calling. The filters were kept as similar as possible to maximize combinability. For AD hybrids, a combined reference of H99 (Janbon *et al.* 2014) and JEC21 (Loftus *et al.* 2005) was used for alignment and SNP identification.

To examine variations in ploidy across the genome, the depth of BWA alignments at all positions was computed using

Samtools mpileup, and then the average depth computed for 5-kb windows across the genome.

MAT locus determination

To evaluate the mating-type alleles present in each isolate, Illumina reads were aligned using BWA-MEM to a Multi-FASTA of both versions of the mating-type locus [AF542529.2 and AF542528.2 (Lengeler *et al.* 2000b)]. Depth at all positions was computed using Samtools mpileup, and then the average depth computed for the *SXI* and *STE20* genes for both idiomorphs. Nearly all isolates showed unique mapping to either the *MAT***a** or *MAT*α alleles of both genes; one isolate, Ftc158, showed significant mapping to both *MAT***a** and *MAT*α, though twofold more to *MAT*α. For the hybrid haploid isolates, the ancestry of the *MAT* locus was determined from the Structure site-by-site output.

Genome assembly and annotation

Illumina sequences for each isolate were assembled using ALLPATHS (Maccallum *et al.* 2009) for 36 isolates (see Table S4 for release numbers for each assembly), or SPAdes 3.6.0 (Bankevich *et al.* 2012) (with parameter –careful) for the remaining three isolates. Assemblies with both fragment and jump libraries were more contiguous than those with fragment-only data (average of 84 or 561 scaffolds, respectively; Table S4). However, there was little difference in the total contig length between assemblies with or without jump data (average 18.4 and 18.5 Mb, respectively; Table S4).

The predicted protein coding gene set for each assembly was generated by combining three primary lines of evidence. Genes were transferred to each new assembly from the well-annotated H99 assembly (Janbon *et al.* 2014) based on whole-genome nucleotide alignments from nucmer. GeneMark-ES (Ter-Hovhannisyan *et al.* 2008) was run on each assembly to generate a *de novo* set of calls. These two sets were combined and improved using PASA (Haas *et al.* 2008) with RNA-sequencing data of three *in vitro* conditions (YPD, limited media, and pigeon guano) generated for H99 (Janbon *et al.* 2014) and for the VNB isolate Bt85 also input. Repetitive elements were removed from the gene set based on TransposonPSI (http://transposonpsi.sourceforge.net/) alignments or PFAM domains found only in transposable elements. The filtered set was assigned sequential locus identifiers across each scaffold. The average number of 6944 predicted genes across all assemblies (Table S4) is close to the 6962 predicted on the H99 reference.

Ortholog identification and comparison

To identify orthologs across the set of 45 *Cryptococcus* genomes (Table S4), proteins clustered based on BLASTP pairwise matches with expect $<1e^{-5}$ using OrthoMCL version 1.4 (Li *et al.* 2003). To identify orthologs specific to each of the serotype A lineages, we required that genes were present in 90% of the assembled genomes for VNI (36 or more) or VNB (eight or more), or all VNII (three genomes). To confirm that orthologs were missing in the other two lineages, synteny was examined around each gene; in some cases this identified candidate orthologs missed by OrthoMCL, which were confirmed by BLASTP similarity and removed.

Phylogenetic analysis

A phylogeny for the sets of 159 or 164 isolates was inferred from SNP data using RAxML version 8.2.4 (Stamatakis 2014) with model GTRCAT and 1000 bootstrap replicates. A separate analysis of the phylogenetic relationship based on gene content included 40 *C. neoformans* var. *grubii* serotype A genomes (28 VNI, 3 VNII, and 9 VNB), 1 *C. neoformans* var. *neoformans* serotype D genome (JEC21), and 4 *C. gattii* genomes (WM276, R265, CA1873, and IND107) (Table S4). The 4616 single-copy orthologs identified in all genomes were aligned individually with MUSCLE (Edgar 2004) at the protein level, converted to the corresponding nucleotide sequence to maintain reading-frame alignment, poorly aligning regions were removed using trimAl (Capella-Gutiérrez *et al.* 2009), and invariant sites were removed. A phylogeny was inferred using RAxML version 7.7.8 in rapid bootstrapping mode with model GTRCAT and 1000 bootstrap replicates.

Population structure

To examine major population subdivisions, we examined how isolates clustered in a principal components analysis (PCA). SNP calls for all the isolates were compared using smartpca (Patterson *et al.* 2006). To identify the major ancestry subdivisions and their contributions to the isolates appearing at intermediate positions in the PCA, a total of 338,562 randomly subsampled positions containing variants in at least two isolates and <5% missing data were clustered using the Bayesian model-based program Structure version 2.3 (Pritchard *et al.* 2000) in the site-by-site mode. Ancestry was plotted across the genome for each isolate using the maplotlib plotting package in Python.

For analysis of *C. neoformans* var. *grubii* diploid isolates (Table S3), diagnostic SNPs for VNB and VNII were present exclusively in the respective groups, and called for all VNB, VNII, and ≥100 VNI isolates. Diagnostic SNPs for VNI were present exclusively in VNII and VNB, and called for all VNB, VNII, and ≥100 VNI isolates.

Population genetic measures including nucleotide diversity (π), fixation index (F_{ST}), and Tajima's D were calculated using popGenome (Pfeifer *et al.* 2014). d_N and d_S measures were calculated from fixed SNPs in each lineage using codeml version 4.9c (Yang 2007). To examine the distribution of the alleles within VNB, we first identified 445,193 alleles private to VNB (present in at least one VNB isolate and no VNI or VNII isolates). We subdivided VNB into four clades (VNBI-South America, VNBI-Africa, VNBII-South America, and VNBII-Africa) and calculated the number of those private alleles unique to each clade (present in that one clade and no others) and shared across VNB groups or geography (present in the two compared clades but no others). The Mantel test was conducted using the center point of each country to determine distances between isolates and the number of

Table 1 Properties of sequenced isolates

Population	Isolates (no.)	*MAT*α	*MAT***a**	*MAT*α/*MAT*α	*MAT***a**/*MAT***a**	*MAT***a**/*MAT*α
Haploid isolates						
VNI	111	109	2			
VNII	23	23	0			
VNB	25	21	4			
VNI–VNB	5	1	4			
VNII–VNB	2	2	0			
Diploid isolates						
VNI–VNB	1			1	0	0
VNII–VNB	2			2	0	0
VNB–Cnn	8			0	1	7
VNB–Cg	1			1	0	0

For each population, the total number of isolates analyzed and the mating type(s) of the isolates are given. Cnn, *C. neoformans* var. *neoformans*; Cg, *C. gattii*.

SNPs between each pairwise set of isolates. The test was conducted using available Python software (https://github.com/jwcarr/MantelTest) with 1000 permutations and the upper tail test of positive correlation.

Linkage disequilibrium

Linkage disequilibrium (LD) was calculated in 500-bp windows of all chromosomes, except for the ~100-kb mating-type locus on chromosome 5, with VCFtools version 1.14 (Danecek *et al.* 2011) using the –hap-r2 option with a minimum minor allele frequency of 0.1.

Population inference by fineSTRUCTURE

Model-based clustering by fineSTRUCTURE (Lawson *et al.* 2012) assigns individuals to populations based on a coancestry matrix created from SNP data, using either Markov chain Monte Carlo or stochastic optimization. The algorithm uses chromosome painting, which is an efficient way of identifying important haplotype information from dense data such as SNP data, and efficiently describes shared ancestry within a recombining population. Each individual is painted using all the other individuals as donors. For example, if an isolate *x* is clonal and a donor, the clonally related recipients will receive almost all of their genetic material from isolate *x*, and its closest relatives. This approach has been applied to analyze recombination in fungal (Engelthaler *et al.* 2014) and bacterial studies (Yahara *et al.* 2013).

fineSTRUCTURE analysis (Lawson *et al.* 2012) was performed using an all-lineage SNP matrix, with one representative of each clonal VNI population to infer recombination, population structure, and ancestral relationships of all lineages. A separate analysis of all VNI lineage isolates was also performed. This approach was based on the presence or absence of shared genomic haplotypes. ChromoPainter reduced the SNP matrix to a pairwise similarity matrix under the linked model, which uses information on LD, thus reducing the within-population variance of the coancestry matrix relative to the between-population variance. Since the *MAT* idiotypes introduce large bias into SNP analysis, they were removed to enable characterization of more defined populations. There was no significant loss of sharing of genetic material when compared to retaining the *MAT* locus.

Data availability

All sequence data from this study have been submitted to GenBank under BioProject identification no. PRJNA384983 (http://www.ncbi.nlm.nih.gov/bioproject); individual accession numbers are listed in Table S1 and Table S4.

Results

Population subdivisions and detection of genetic hybrids

To examine the evolution of *C. neoformans* var. *grubii*, we sampled the population by sequencing the genomes of 188 isolates (Table 1 and Table S1) representing each of the three major genetic subpopulations (VNI, VNII, and VNB) previously defined using multi-locus sequence typing (MLST) (Litvintseva *et al.* 2006; Meyer *et al.* 2009). These isolates are geographically diverse, originating from North America, South America, the Caribbean, Asia, Europe, and Africa (Table S1). The VNI global lineage is the most geographically diverse, whereas VNII is represented by a smaller number of locations and VNB appears most highly prevalent in southern Africa. For VNI and VNB, both clinical and environmental isolates were included, with 25 VNI isolates originating from avian guano or trees and 8 VNB isolates from trees or other environmental sources (Table S1). For each isolate, we identified SNPs using GATK by aligning Illumina reads to the H99 reference genome assembly (*Materials and Methods*; Janbon *et al.* 2014). Whereas 164 isolates appeared haploid, 24 isolates were determined to be heterozygous diploids (*Materials and Methods*, Table 1) and analyzed separately. An initial phylogeny of the 164 haploid isolates separated the three lineages, but intermediate placement of five isolates suggested the presence of hybrid haploid genotypes (Figure S1). As the phylogenetic placement of such hybrid isolates is complicated by recombination, we removed these isolates from the phylogenetic analysis and analyzed them using alternative approaches (see below).

A phylogeny inferred from the SNPs for all nonhybrid isolates strongly supports the three major lineages of *C. neoformans* var. *grubii*: VNI, VNII, and VNB (Figure 1). Of these 159 isolates, only six (4%) contain the rare *MAT***a** allele,

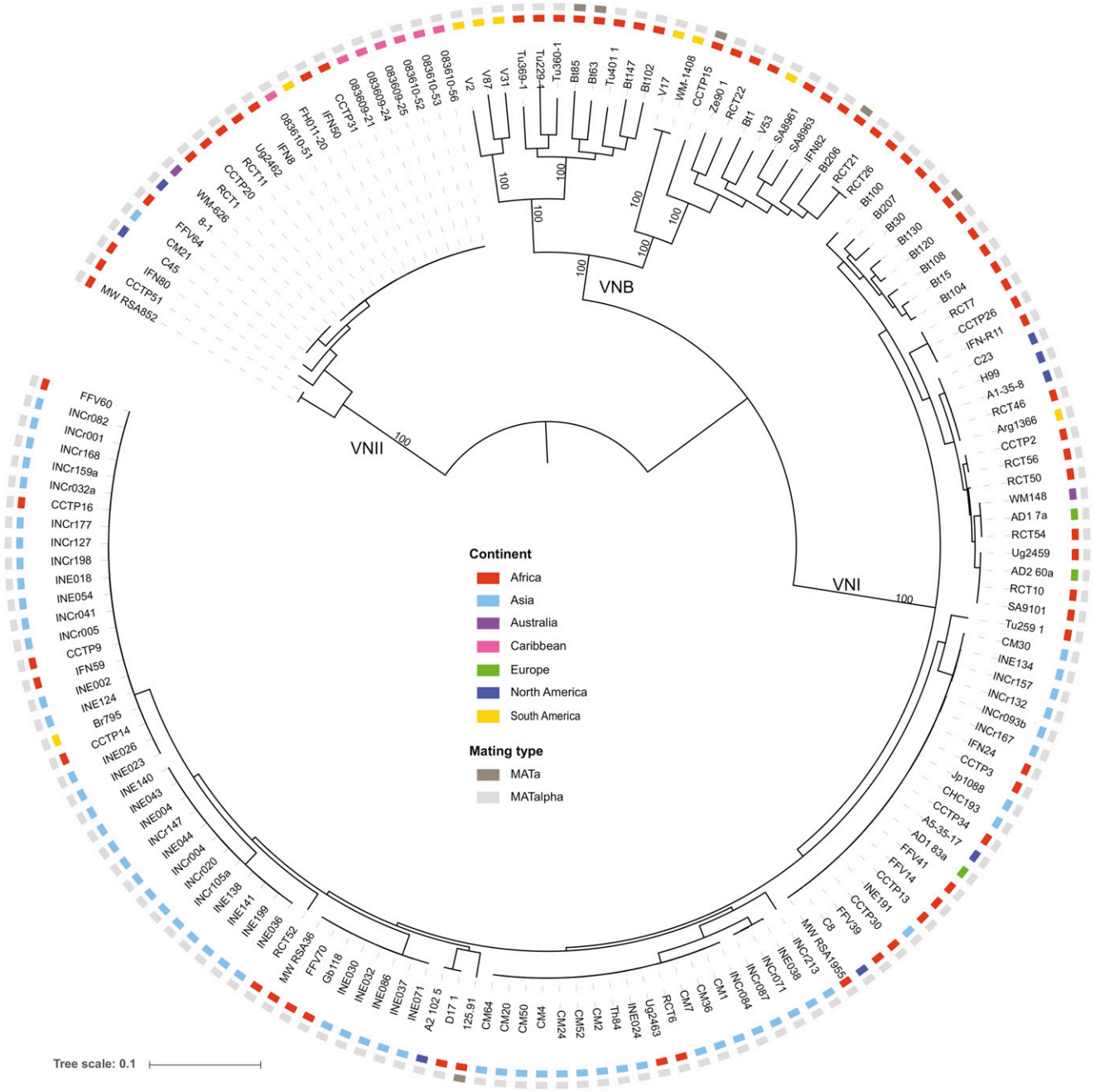

Figure 1 Phylogenetic analysis supports three major lineages of *C. neoformans* var. *grubii*. Using a set of 876,121 SNPs across the 159 nonhybrid isolates, a phylogenetic tree was inferred using RAxML. The tree was rooted with VNII as the outgroup (Hagen *et al.* 2015). The percentage of 1000 bootstrap isolates that support each node is shown for major nodes with at least 90% support. For each isolate, the geographic site of isolation is noted by colored boxes.

including four VNB isolates (Bt63, Bt85, Bt206, and CCTP15) and two VNI isolates (125.91 and Bt130). Based on these whole-genome SNP comparisons, none of these *MAT***a** isolates appeared highly related to each other or to any *MAT*α isolate. The two VNI *MAT***a** isolates are well separated within this group, with Bt130 found in a subgroup of African isolates and 125.91 most closely related to a pair of isolates from Africa and North America (Figure 1). Phylogenetic analysis showed that VNB has the highest diversity between isolates, showing the longest tip branches compared to VNI or VNII. In addition, VNB consisted of two diverged subgroups, VNBI and VNBII, as suggested previously by MLST (Litvintseva *et al.* 2006, 2011; Chen *et al.* 2015) and genomic analysis (Desjardins *et al.* 2017; Vanhove *et al.* 2017).

To better understand the population structure of the three lineages and identify potential interlineage recombination, we compared results of two independent approaches. First, we used PCA to identify the major groups in the population using

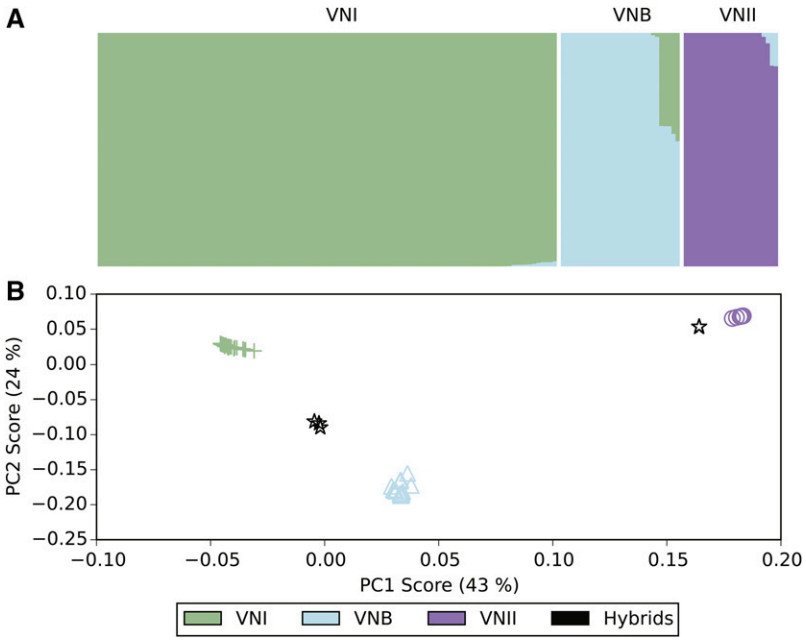

Figure 2 Ancestry characterization of three major groups highlights hybrid isolates. (A) The fraction of ancestry (k = 3) estimated by Structure is shown within a column for each isolate. (B) PCA separates the three major lineages, with the hybrid isolates showing a mix of VNB ancestry with either VNI or VNII.

the SNP data. By comparing the SNP variants across isolates using PCA, we found there are three major clusters corresponding to the VNI, VNII, and VNB lineages (Figure 2). The five isolates that showed intermediate positions in phylogenetic analysis (Figure S1) also appeared at intermediate positions by PCA, placed between VNI and VNB. In addition, two isolates were separated from the VNII cluster and shifted toward the VNB cluster. All of these seven isolates were collected from southern Africa, and all had a clinical origin except isolate Ftc260-1, which was isolated from the environment (Table S1). Of the seven, two sets of isolates share nearly identical ancestry ratios and appear closely related on the phylogenetic tree. Isolates Bt131, Bt162, and Bt163 differed by an average of only 39 SNP positions; similarly, CCTP51 and MW_RSA852 differed by 200 SNP positions, suggesting these five isolates are descended from two hybridization events. Therefore, four unique hybridization events were detected in total, three for VNI–VNB and one for VNII–VNB. While the basal branching VNB isolates from Brazil could suggest a hybrid ancestry, all appear to be uniformly VNB (>99% of sites).

Next, we identified the ancestry contribution of each isolate using Structure with three population subdivisions. This confirmed that most isolates have a single dominant ancestry assigned to the VNI, VNB, and VNII lineages. In addition, the isolates with intermediate positions indicated by PCA were found to have mixed ancestry contributions by Structure. SNP sites for the VNI–VNB hybrids contain an average of 40.8% VNI ancestry and 59.2% VNB ancestry, whereas the VNII–VNB hybrids have 85.8% VNII and 14.2% VNB ancestry (Table S2). The similar fraction of ancestry in the VNI–VNB hybrids suggests they could be recent mixtures of the two lineages, whereas the VNII–VNB hybrids may be more ancient mixtures with additional crosses to VNII isolates biasing the final ratio of parental SNPs.

Evidence of recent meiotic exchange generating haploid hybrids

To examine the degree of intermixing of ancestry for these hybrid genotypes across the genome, we identified the most-likely ancestry for each SNP site using the site-by-site mode in Structure. Selecting positions where the ancestry assignment was most confident (≥0.9; *Materials and Methods*), we examined the distribution of these sites by ancestry across the 14 chromosomes (Figure 3). Each of the three VNI–VNB hybrids displayed different patterns of large regions corresponding to a single ancestry. For example, chromosome 1 has three large blocks of different ancestry in Bt125, four in Bt131, and two in Ftc260-1 (Figure 3, A–C). While all chromosomes contained regions of both VNI and VNB ancestry groups in Bt125 and Ftc260-1, two chromosomes of Bt131, chromosome 6 and 9, have only large regions of VNB ancestry. By contrast, CCTP51, which contains a lower fraction of the second ancestry (VNB), appears more highly intermixed with smaller ancestry blocks (Figure 3D). Notably, three of the four unique genotypes (Bt131, CCTP51, and Ftc260-1) contain the rare *MAT***a** locus; in all *MAT***a** isolates, the mating-type locus region is of VNB ancestry, whereas the mating locus region in the *MAT*α isolate (Bt125) is of VNI ancestry (*Materials and Methods*). Overall, these patterns suggest a recent hybridization of VNI and VNB isolates, with recombination during meiosis generating chromosome-wide intermixing, resulting in distinct parental haplotype blocks. In Bt125, a 205-kb region of scaffold 6 is present at nearly twice (1.92-fold) the average depth. Otherwise, this isolate and the other six hybrid isolates were found to contain even levels of ploidy across the 14 chromosomes based on read depth.

For the three VNI–VNB hybrids showing large ancestry blocks, we also used the site ancestry predictions to finely map the genotypes within each population. Given the roughly equal contribution of the two ancestry sites and the

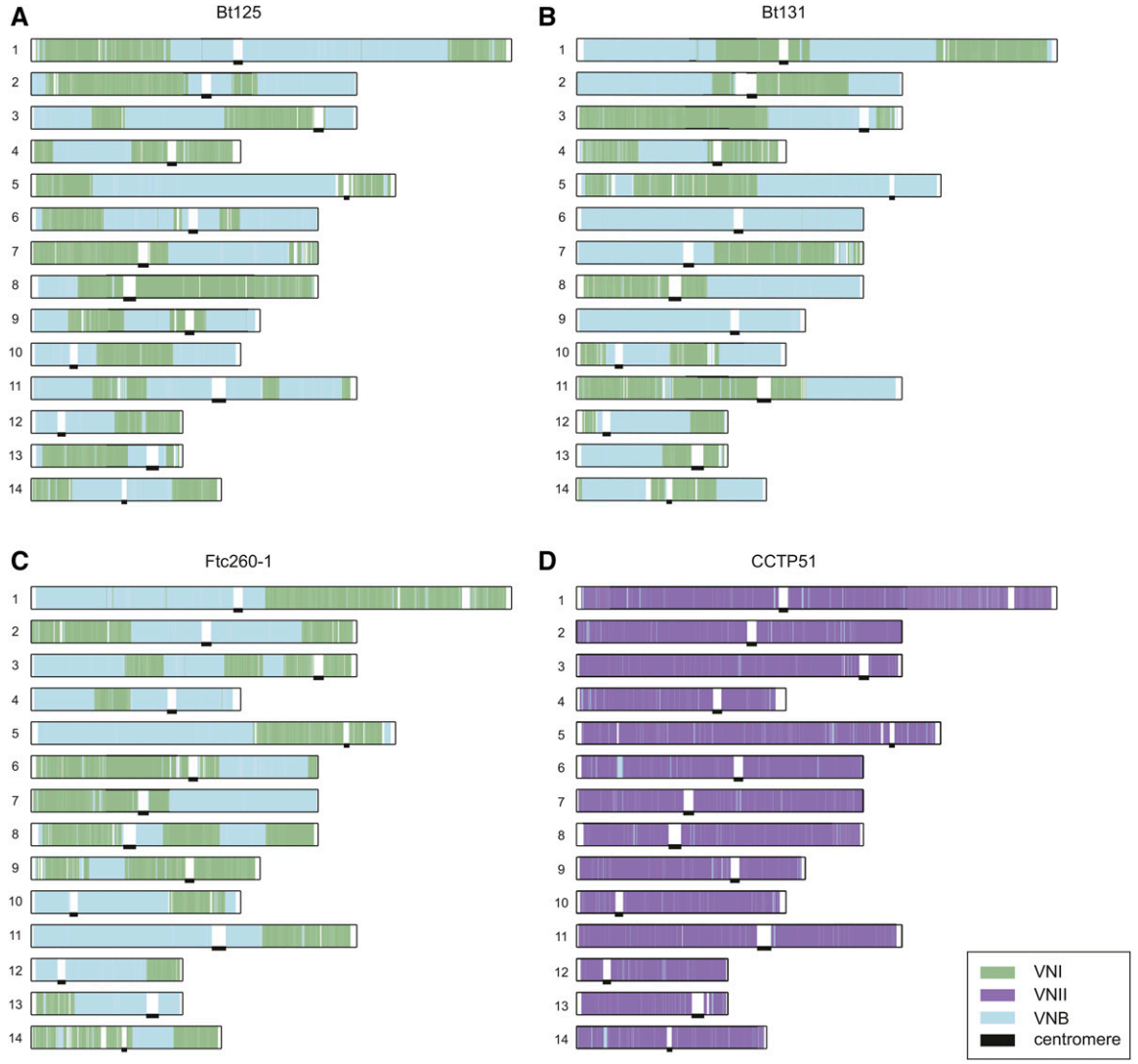

Figure 3 Large blocks of ancestry suggest recent recombination between lineages. For each of the four isolates depicted (A, Bt125; B, Bt131; C, Ftc260-1, and D, CCTP51), the Structure-assigned ancestry for each site along each chromosome is depicted as a colored bar corresponding to VNI, VNII, and VNB ancestry. Locations of centromeres are marked with black bars.

large block size for each in these genomes, we hypothesized that these hybrids could have resulted from recent mating of one genotype of each lineage, which we could reconstruct using separate phylogenies of each site class. For each genotype, sites mapped to either the VNI or VNB ancestry were selected and a separate phylogeny constructed for each of these two sets of sites. For VNI ancestry sites, these isolates had very different genotypes, with Ftc260-1 most closely related to a diverse set of African isolates in VNI; whereas both Bt125 and Bt131 are more closely related to highly clonal clades of VNI isolates (Figure S2, A, C, and E). Similarly for a separate phylogenetic analysis of VNB ancestry sites, Bt125 and Bt131 were placed within the VNBII subclade of VNB, while Ftc260-1 was placed in VNBI (Figure S2, B, D, and F). This supports that these three hybrids originated from very different genotypes of VNI and VNB parental isolates.

Diploid isolates and genome plasticity

As noted above, a total of 24 sequenced isolates displayed heterozygous SNP positions across the genome. Four of these isolates had higher rates of polymorphism overall and appear to be hybrids within or between VN lineages (Bt66, Cng9, PMHc.1045.ENR.STOR, and 102-14) (Figure S3). Each of these isolates contain two copies of the *MAT*α mating-type locus which show similar levels of heterozygosity as the rest of the genome, suggesting that these diploids arose from same-sex mating of two *MAT*α parental isolates with different genotypes. In addition, 11 serotype-A diploids showed very low rates of heterozygosity (Figure S3), consistent with AFLP- and MLST-based evidence that they arose from endoreduplication or self-mating (Lin *et al.* 2009). The remaining isolates include eight serotypeA/serotypeD diploids, of which seven contain both *MAT***a** and *MAT*α mating types and one is

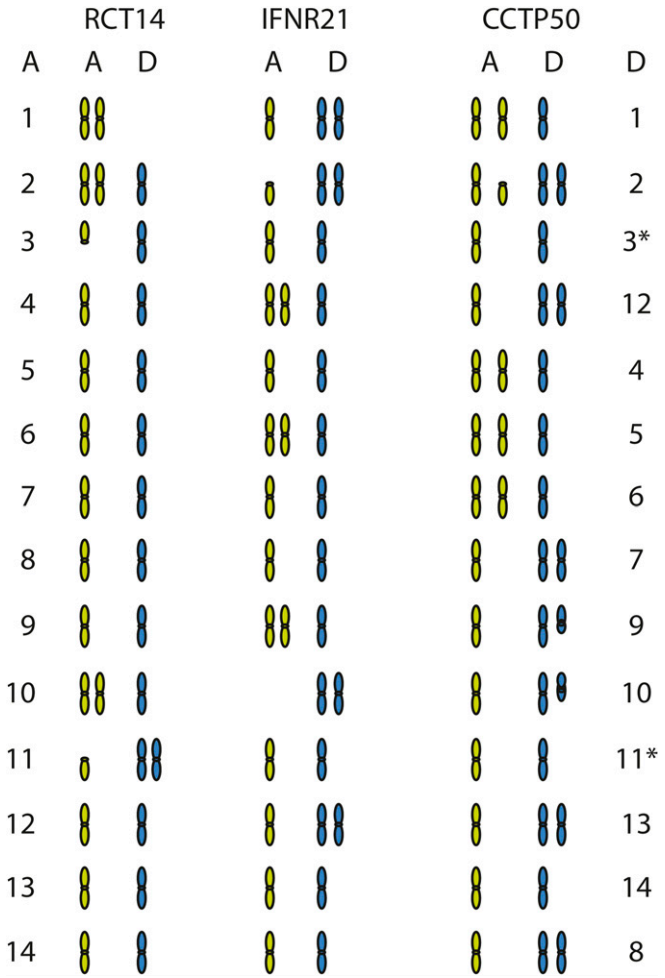

Figure 4 Chromosome ancestry and ploidy variation of AD hybrids. For three AD hybrid isolates (RCT14, IFNR21, and CCTP50), the contribution and copy number of A (green) and D (blue) ancestry chromosomal regions was measured by aligning all sequence reads to a combined AD reference (A: H99, left, and D: JEC21, right). The copy number of each chromosome is depicted, with either full or partial chromosomal regions shown; see Figure S4 for detailed coverage plots for all AD hybrid isolates.

homozygous for the *MAT***a** locus, and one serotype A/*C. gattii* hybrid contains two copies of *MAT*α.

All types of diploid isolates in our set, including AA diploids, exhibit regions of loss of heterozygosity (LOH) in the genome, where alleles of only one parental isolate are present. Three of the AA diploids (Bt66, Cng9, and 102-14) are heterozygous throughout nearly all of the genome; Cng9 exhibited only a small LOH region at the start of chromosome 2, which also has haploid levels of genome coverage. Isolate PMHc1045, by contrast, has large LOH regions on six scaffolds, including a 1.1-Mb region of chromosome 6 (Figure S3). Some of these regions of LOH in PMHc1045 are linked to aneuploid chromosome segments, including a region of chromosome 12 reduced to haploid levels and/or triploid levels of the region adjacent to an LOH on chromosome 6. All LOH regions are telomere linked, reminiscent of what has previously been reported across diverse isolates of *Candida albicans* (Hirakawa *et al.* 2015).

We next inferred the ancestry of the two parental isolates contributing to the AA hybrids by examining the frequency of SNP alleles that are highly predictive for VNI, VNII, or VNB (*Materials and Methods*). Three of the isolates (Cng9, PMHc1045, and 102-14) have similar frequencies of such VNII and VNB alleles, whereas Bt66 is comprised of VNI and VNB predictive alleles (Table S3). Comparing Cng9 and PMHc1045 directly, 89.2% of variant sites are identical; this fraction increases to 97.3% when LOH regions are excluded and a similar fraction of sites are shared with 102-14. Notably, LOH has resulted in a mixing of genotypes: examining predictive alleles for each of the seven LOH regions of PMHc1045 (Figure S3) revealed two regions encompassing 1.4% of the genome share the highest fraction of private alleles with other VNB isolates, whereas the remaining five regions encompassing 10.2% of the genome share most private alleles with other VNII isolates. By contrast, Cng9 has only a single small region of LOH that does not overlap with any of the seven LOH regions in PMHc1045. Thus, LOH has led to large differences between otherwise highly similar Cng9 and PMHc1045 isolates and resulted in blended ancestry by converting regions to each of the two parents in PMHc1045.

The eight AD hybrids also showed evidence of even more extreme aneuploidy and LOH related to loss of one of the two parental chromosomes. All isolates displayed evidence of aneuploidy when examining read coverage across both the H99 serotype A and JEC21 serotype D reference genomes (Figure S4). While some isolates have retained chromosomes of both A and D origin, others have lost a chromosome from one parent and duplicated the corresponding chromosome of the other (Figure 4 and Figure S4). For example, in RCT14, two copies of chromosome 1 are present but both have serotype A origin; similarly in IFNR21, both copies of chromosome 10 have serotype D origin. Both of these isolates display additional aneuploidies, with three copies of some chromosomes. Notably, CCTP50 appears mostly triploid, with A/D ratios of either 2:1 or 1:2 for each chromosome (Figure 4); this pattern is also observed in IFN26 (Figure S4). In IFN-R26, loss of chromosome 4 in JEC21, balanced by gain of chromosome 5 in H99 (Figure S4), has resulted in a *MAT***a**/*MAT***a** genotype. While the mating type of the original JEC21 parent cannot be determined, this suggests that generation of *MAT***a**/*MAT***a** diploids can occur via chromosome loss and duplication. All other isolates are *MAT***a**/*MAT*α, suggesting that they originated from opposite sex mating. While diploid AD hybrids have been isolated from both environmental and clinical sources (Litvintseva *et al.* 2006), all eight AD hybrids in our set are of clinical origin.

To examine the diversity of these AD hybrids, SNPs were identified by comparison to a combined A (H99) and D (JEC21) genome reference. Phylogenetic analysis of A and D genome SNPs revealed that both the A and D copies of each hybrid are closely related for these isolates (Figure S5). On average, the A genomes differ by 6108 SNP positions and the D genomes by 3935 SNP positions. The A genomes are from

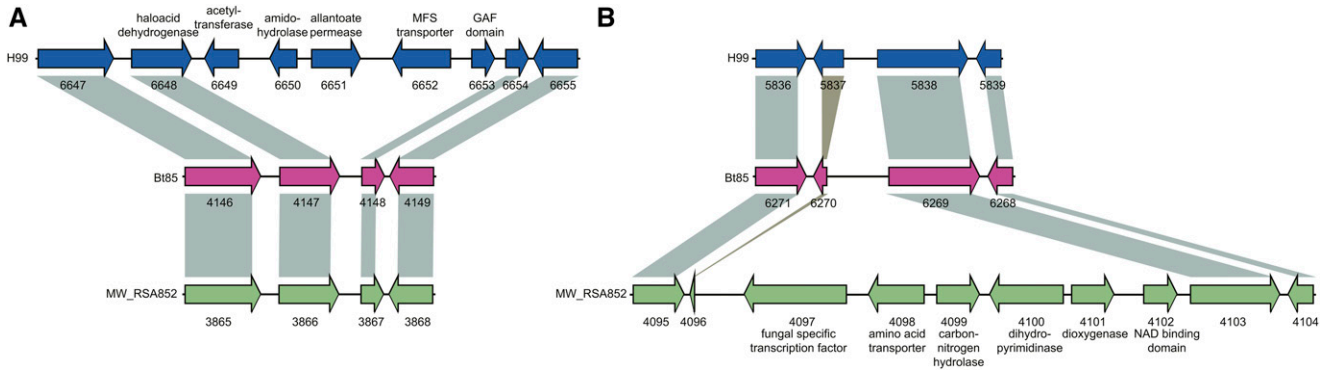

Figure 5 Lineage-specific gene clusters. Two large lineage-specific clusters were detected in the VNI genomes or VNII genomes; these are depicted using a representative genome from each lineage. (A) Insertion of CNAG_06649 to CNAG_06653 in H99 (blue, VNI). Syntenic genes in Bt85 (pink, VNB) and MW_RSA852 (green, VNII) are connected with gray bars. (B) Insertion of C358_04097 to C358_04102 in MW_RSA852.

the VNB lineage, most closely related to Bt206 in our analysis (Figure S5). The low diversity of both the A and D genomes between isolates suggests that this set of eight AD hybrids may have originated from a single hybrid isolate or from a set of closely related A and D parental isolates.

Chromosomal copy number variation

On a smaller scale than whole-genome hybridization, chromosomal copy number variants appear to be common in *C. neoformans* and may be an adaptive mechanism for virulence (Rhodes *et al.* 2017). In the set of 164 primarily haploid isolates, 25 exhibited whole or partial chromosomal aneuploidies (Figure S6). In 13 of the 25 isolates, an entire chromosome or region thereof showed a doubling of sequencing coverage, consistent with a diploid chromosome in an otherwise haploid isolate. The remaining 12 isolates showed a 50% gain in coverage better explained by a diploid isolate with a triploid chromosome or region. These likely diploid isolates do not display heterozygous base calls, suggesting a recent endoreduplication of the genome and associated aneuploidy of additional chromosomes.

Aneuploidies of particular chromosomes may provide a specific biological advantage or alternatively be better tolerated. In general, the smallest chromosomes (12 and 13) are the most frequently observed to exhibit aneuploidy (Figure S6). Several isolates have an increased copy number of chromosome 1; amplification of the lanosterol-14-α-demethylase *ERG11* and the major efflux transporter *AFR1* located on chromosome 1 can confer resistance to azole drugs (Sionov *et al.* 2010). Of the four isolates that contain chromosome 1 aneuploidies, either *ERG11* (CCTP34) or *AFR1* (IFN-R11 and RCT6) or both genes (CCTP9) are present at elevated copy number. The elevated copy number of *AFR1* appears correlated with increased drug resistance; both CCTP9 and RCT6 displayed fluconazole MIC values of 256 μg/ml, whereas CCTP34 appeared more susceptible at an MIC of 8 μg/ml (*Materials and Methods*). Notably, all of the isolates with chromosome 1 aneuploidies are of clinical origin, as are 24 of all 25 isolates with detected aneuploidies (Figure S6 and Table S1). Of the seven isolates with hybrid ancestry, only Bt125 included a small region of chromosome 6 at higher copy number; otherwise, this and the other hybrid isolates appeared to be haploid. Across the diploid and haploid isolates, we detected aneuploidies affecting all chromosomes (Figure S3, Figure S4, and Figure S6).

Conservation of gene content and structure across lineages

To examine the extent of gene content variation across the three major lineages of *C. neoformans* var. *grubii*, we assembled and annotated genomes of 39 representative isolates (*Materials and Methods*). Previously, a high quality reference genome was produced for the H99 VNI isolate (Janbon *et al.* 2014); our data set includes new annotated assemblies for 9 diverse VNB isolates, 27 VNI isolates, and 3 VNII isolates (Table S4). The gene sets across all 40 assemblies (including H99) were compared to each other and to those of four *C. gattii* (representing VGI, VGII, VGIII, and VGIV) and one *C. neoformans* var. *neoformans* (serotype D) reference genomes (*Materials and Methods*) to evaluate gene conservation. Based on orthologs identified across these genomes (*Materials and Methods*), an average of 4970 genes are conserved across all 45 compared *Cryptococcus* gene sets; within serotype A, an average of 5950 genes are conserved in all 40 genomes (Figure S7). A phylogeny inferred from 4616 single copy genes supports VNII in an ancestral position relative to the more recently diverging VNI and VNB (Figure S7; 100% bootstrap support), solidifying results previously seen with targeted sequencing of 11 nuclear loci (Hagen *et al.* 2015).

Gene content is highly conserved across *C. neoformans* var. *grubii* with few examples of genes specific to the separate lineages (File S1). Based on ortholog profiling, a total of 11 genes are specific to VNI, 3 specific to VNB, and 25 specific to VNII (Table S5); by comparison 59 genes are conserved in C. neoformans var. grubii but absent in C. neoformans var. neoformans and C. gattii (Table S6, File S1). These include two clusters of genes specific to VNI or VNII located within otherwise syntenic regions of the genome (Figure 5). The cluster of five genes unique to VNI genomes include a predicted haloacid dehydrogenase, an amidohydrolase, and an

Table 2 Rapidly evolving genes in the three lineages of *C. neoformans* var. *grubii*

Comparison	d_N	Locus	Gene	Annotation
VNI *vs.* VNB	0.0181	CNAG_01841	*GLN3*	Transcription factor, deletion sensitive to organic peroxides (Jung *et al.* 2015)
	0.0155	CNAG_03894	*PDR802*	Transcription factor, deletion with reduced virulence (Jung *et al.* 2015)
	0.0095	CNAG_03213	*UVE1*	UV damage endonuclease
	0.0092	CNAG_02756	*CDC43*	Geranylgeranyltransferase-I, essential for virulence (Selvig *et al.* 2013)
	0.0090	CNAG_06655	*GPI18*	GPI-anchor transamidase
	0.0089	CNAG_01908	*HEM4*	Uroporphyrinogen-III synthase
	0.0085	CNAG_03133	*ATG2602*	UDP-glucose sterol transferase
	0.0084	CNAG_03617	*CLP1*	Clampless protein 1
	0.0076	CNAG_05740	*RAM1*	Farnesyltransferase β-subunit, essential for virulence (Esher *et al.* 2016)
	0.0068	CNAG_03637	*YKU80*	Double-strand break repair factor and silencing regulator, deletion has reduced virulence (Liu *et al.* 2008)
VNI *vs.* VNII	0.0610	CNAG_05836	*HOC1*	α-1,6-mannosyltransferase (Lee *et al.* 2015)
	0.0408	CNAG_05838	*RGD1*	Rho GTPase activating protein, deletion has increased virulence (Liu *et al.* 2008)
	0.0214	CNAG_06031	*KRE63*	β-glucan synthase, involved in capsule and cell wall formation, deletion has decreased virulence (Gilbert *et al.* 2010)
	0.0149	CNAG_06814	*SXI1α*	α cell type transcription factor, required for mating (Hull *et al.* 2002)
	0.0142	CNAG_01841	*GLN3*	See above
	0.0135	CNAG_03229	*YOX101*	Transcription factor, deletion sensitive to organic peroxides (Jung *et al.* 2015)
	0.0127	CNAG_03398	*ZIP2*	Zinc ion transporter
	0.0113	CNAG_03133	*ATG2602*	See above
	0.0110	CNAG_03366	*ZNF2*	Transcription factor, overexpression results in reduced virulence (Wang *et al.* 2012)
	0.0104	CNAG_01019	*SOD1*	Superoxide dismutase
VNB *vs.* VNII	0.0617	CNAG_05836	*HOC1*	See above
	0.0402	CNAG_05838	*RGD1*	See above
	0.0171	CNAG_06031	*KRE63*	See above
	0.0128	CNAG_03366	*ZNF2*	See above
	0.0122	CNAG_06814	*SXI1α*	See above
	0.0114	CNAG_03213	*UVE1*	See above
	0.0104	CNAG_01019	*SOD1*	See above
	0.0104	CNAG_03398	*ZIP2*	See above
	0.0102	CNAG_01841	*GLN3*	See above
	0.0102	CNAG_02756	*CDC43*	See above

Consensus sequences were built for each lineage, and d_N and d_S were calculated for each lineage pair. As d_S was uniformly low throughout the data set due to limited genetic diversity, for each pair of lineages we identified the 10 genes with assigned names (Inglis *et al.* 2014) with the highest d_N, which measures both the mutation rate and selection.

allantoate permease, which could be involved in uptake of uric acid products. The cluster of six genes unique to the VNII genomes includes a predicted transcription factor, amino acid transporter, hydrolase, dihydropyrimidinase, and oxygenase superfamily protein. While both clusters are also missing from the JEC21 *C. neoformans* var. *neoformans* genome, the more distantly related *C. gattii* genomes contain syntenic orthologs of all of the VNII-specific cluster genes and between one and three nonsyntenic orthologs of the VNI-specific cluster. These patterns suggest gene loss and perhaps lateral transfer in some species, and the lineages account for these differences. There was little other evidence of lineage-specific gene loss; orthologs missing in only one lineage included only hypothetical proteins. In addition, we further searched for genes with loss-of-function mutations in all members of each lineage, using SNP data, to find genes that may be disrupted but still predicted in the assemblies. However, we found no convincing evidence of disrupted genes with known functions in all members of any of the three lineages (File S1).

Given the high level of gene conservation between lineages, we sought to identify rapidly evolving genes that might be involved in phenotypic differences between *C. neoformans* lineages. For each gene, we built a consensus sequence for each lineage and then calculated pairwise d_N and d_S of these fixed sites. As d_S was uniformly low throughout the data set due to limited genetic diversity, we identified differences in d_N, which measures both the mutation rate and selection. The top 10 annotated genes with the largest d_N for each pairwise comparison are shown in Table 2, and the three comparisons in total include 18 unique genes. The set is dominated by transcription factors (*GLN3*, *PDR802*, *SXI1α*, *YOX101*, and *ZNF2*) and transferases (*ATG2602*, *CDC43*, *GPI18*, *HOC1*, and *RAM1*), many of which have already been implicated in virulence (Wang *et al.* 2012; Selvig *et al.* 2013; Jung *et al.* 2015; Lee *et al.* 2015; Esher *et al.* 2016) or resistance to oxidative stress (Jung *et al.* 2015). In particular, *CDC43* and *RAM1* are both rapidly evolving; these genes represent the two major independent methods of prenylation, which is the key to proper subcellular localization of many proteins, often to the membrane (Selvig *et al.* 2013; Esher *et al.* 2016). Other rapidly evolving genes include β-glucan synthase *KRE63*, superoxide dismutase *SOD1*, and mating regulator *SXI1α*, the latter of which is highly divergent between VNII

Table 3 Population genetic features of the lineages of *C. neoformans* var. *grubii*

Populations	Isolates (no.)	Segregating sites	π	Tajima's D
VNI	111	190,716	0.00200	−0.107179
VNII	23	337,990	0.00105	−1.005950
VNB	25	613,991	0.00736	−0.232596

The total number of isolates, number of segregating sites, π, and Tajima's D are given for each population.

and both VNI and VNB, and could play a role in reproductive isolation of the VNII lineage.

Population measures and biogeography

Strikingly, recently identified VNB genotypes from South America are placed in the phylogeny as basally branching clades for each VNB subgroup, which otherwise consist of genotypes from Africa (Figure 1). All of the six South American VNB isolates contain the *MAT*α genotype. By contrast, both VNI and VNII consist of more closely related, though more geographically diverse, sets of isolates; one large clonal group is found in VNII, whereas several are observed for VNI, which is oversampled owing to its higher prevalence in patients and environments worldwide. Overall, VNB showed the highest average pairwise diversity ($\pi = 0.00736$), nearly four times the level in VNI ($\pi = 0.00200$), with the lowest value for VNII ($\pi = 0.00105$) (Table 3). Genetic diversity within the VNB lineage was similar between the South American and African isolates ($\pi = 0.00727$ and 0.00736, respectively). However, genetic diversity of VNI isolates in India was lower than VNI isolates in Africa ($\pi = 0.00146$ and 0.00337). VNB also contained the largest fraction of private alleles compared to VNI and VNII, reflecting the higher variation within VNB (Table 4). By contrast, VNI and VNII had the highest number of fixed differences, reflecting the long branches leading to these clades. The average divergence (dXY) between the lineages' ranges is 0.012 for comparison of isolates from VNI and VNB, and 0.015 for comparison of either lineage to VNII (Table 4); highlighting the low nucleotide divergence between the lineages. VNI and VNII were the most differentiated of the three lineages as shown by pairwise whole genome fixation indexes (F_{ST}) (Weir and Cockerham 1984). The highest average chromosome F_{ST} value is 0.874 between VNI and VNII isolates, while the average chromosome F_{ST} values of VNI–VNB and VNB–VNII are 0.595 and 0.707, respectively (Table 4).

To further examine the evolutionary history of the novel South American VNB isolates, we subdivided VNB into four subclades (VNBI-South America, VNBI-Africa, VNBII-South America, and VNBII-Africa) and calculated alleles unique to each subclade and shared across VNB groups or geography (*Materials and Methods*). These subclades represent all combinations of the two previously identified VNB groups (VNBI and VNBII) and the two geographies (South America and Africa). One South American VNB isolate (V53), nested deeply within African isolates on the phylogeny, was excluded from the analysis. Each of the four subclades contained more unique alleles than were shared across either VNB group or geography (Figure 6), suggesting both a high level of genetic diversity within each subclade and some degree of reproductive isolation between them. Furthermore, there was a greater number of unique alleles shared within the VNB groups from different geographic regions than were shared across VNB groups within the same geographic region (Figure 6). This geographically and phylogenetically segregated diversity suggests that multiple ancient migration events occurred between South America and Africa during the diversification of VNB, followed by geographic isolation. In contrast, the VNI and VNII lineages showed a pattern consistent with more rapid current migration, where isolates from different geographic regions in many cases differed by <200 SNPs.

We next evaluated whether VNI and VNB showed a signal of genetic isolation by distance using the Mantel test. In both VNI and VNB, genetic distance was significantly positively correlated with geographic distance ($P = 0.0001$ and $P = 0.042$, respectively). When VNB was separated into VNBI and VNBII, each lineage showed an even stronger signal ($P = 0.0051$ and $P = 0.0009$, respectively), suggesting much of the correlation seen within VNB is representative of isolation within each subclade. Therefore, despite VNB showing signals of more ancient migration while VNI shows signals of recent migration, both demonstrate genetic substructure according to geography.

Recombination between and within lineages

The basal branching of Brazilian VNB isolates revealed in the phylogenetic analysis suggested that South America could be a global center of *C. neoformans* var. *grubii* diversity. To further investigate this hypothesis, and to explore recombination in the context of population structure, we implemented the chromosome-painting approach of fineSTRUCTURE (Lawson *et al.* 2012), which identifies shared genomic regions between individuals and thereby ancestral relationships among individuals and populations. Our linked coancestry model found the highest level of sharing among VNB isolates; in addition, there is evidence of strong haplotype donation from South American VNB isolates (V2, V31, and V87) to all other lineages and continents, suggestive of ancestral recombination (Figure 7). Independent confirmation of ancestry using Structure confirmed that V87 includes primarily VNB ancestry with ~1% VNI alleles (Table S7). Interrogating the chunk counts, which are lengths of DNA shared by a donor to other individuals, and lengths produced by fineSTRUCTURE revealed that the haplotype chunks donated by these "ancestral" isolates were substantially higher than those seen for other isolates, with other African VNB isolates receiving significant chunks and lengths (Bt102, Bt63, Bt85, Tu229-1, Tu360-1, Tu369-1, and Tu401-1) from the South American VNB isolates. Isolate V53 donated less strongly than these three isolates to all lineages. Other South American VNB isolates (WM 1408 and V17) donated strongly to specific lineages: WM 1408 to VNII and VNB, while V17

Table 4 Pairwise population genetic statistics between the lineages of *C. neoformans* var. *grubii*

Comparisons	Fixed	Shared	Private_A	Private_B	dXY	F_{ST}
VNB *vs.* VNI	54,719	52,536	446,566	102,817	IvB: 0.0119	IvB: 0.595
VNB *vs.* VNII	118,329	68,211	405,406	78,444	BvII: 0.0154	BvII: 0.707
VNI *vs.* VNII	188,590	38,501	116,845	83,802	IvII: 0.0152	IvII: 0.874

The number of alleles fixed and shared between the populations, and alleles private to each population are given, along with divergence metrics dXY and F_{ST}.

donated to VNI and VNB. However, these findings for WM 1408 and V17 were not corroborated using Structure. Despite their allocation to separate VNB subpopulations, V2 and V17 (VNBI and VNBII, respectively) donate the most genetic material (when interrogating the chunk counts) to VNI isolates in Africa, India, and Thailand.

Within the VNI lineage, fineSTRUCTURE analysis identified a subset of isolates with a high frequency of haplotype sharing (Figure 8). Notably, a group of African (Tu259-1, 125.91, RCT52, Bt100, Bt207, and Bt30) and Indian (INCr213 and INE071) isolates show strong haplotype donation with many other VNI isolates, suggestive of ancestral recombination events. These isolates are dispersed over four subpopulations within the VNI lineage. Though the geographic distance between these populations should preclude frequent intermixing, these isolates from Africa and India may include a higher fraction of ancestral alleles, leading to a lack of phylogeographic structure among these highly geographically distinct populations.

Finding that ancestral recombination in the VNB lineage contributed to VNI lineage diversity suggested that there could be a signature of admixture LD in these two populations. LD differs between lineages (Figure S8), with VNII LD decaying slowly with physical distance, and manifesting an LD50 (where LD has decayed to half its maximum value) at >150 kb. However, this value may reflect the highly clonal nature and relatively small number of sequenced VNII isolates. LD decay is relatively slow for VNI with an LD50 of 4500 bp, whereas LD decays more rapidly in the VNB lineage, with an LD50 of 1500 bp. When separated into geographic origin of isolation (Figure S8b), LD50 for South American VNB appears greater (>150 kbp) than that seen in African VNB (2000 bp). The slower decay of LD in VNI and VNII relative to VNB may reflect a lower frequency of sexual reproduction owing to the rarity of the *MAT***a** idiomorph and therefore meiotic recombination would have fewer opportunities to break apart LD blocks.

Discussion

This population genomic analysis of *C. neoformans* var. *grubii* has revealed new biogeographic relationships and highlighted a complex history of hybridization events between groups. Analysis of genome-wide variation of 188 geographically diverse isolates greatly increases the resolution of the VNI, VNII, and VNB phylogenetic groups and precisely measures the level of genetic differentiation between isolates within each group and across geographic scales. These data support a much higher diversity of isolates in the VNB group compared to VNI and VNII isolates. Notably, we show that hybridization between these groups can result in genome mixing, suggestive of recent and ongoing meiotic exchange, and introgression of smaller regions between lineages have been identified and appear to perpetuate vertically (Desjardins *et al.* 2017). Therefore, although there is good support for the separation of the groups based on phylogenetic analysis, the measures of intermixing that we observe do not meet the strict requirements for species definition under a Genealogical Concordance Phylogenetic Species Recognition (GCPSR) framework (Taylor *et al.* 2000; Dettman *et al.* 2003). The GCPSR defines phylogenetic species by identifying the transition from genealogical concordance to conflict (reticulate genealogies) as a means of determining the limits of species, a requirement that *C. neoformans var. grubii* does not appear to satisfy owing to ongoing gene flow among the lineages. Similarly, a recent taxonomic proposal to divide the *C. neoformans* and *C. gattii* species complexes into seven monophyletic species did not subdivide *C. neoformans* var. *grubii* into separate species; while VNI, VNII, and VNB were strongly supported clades in a multi-locus phylogeny, coalescent-based approaches did not clearly support these three lineages as separate species (Hagen *et al.* 2015). In addition, the interlineage recombination or hybridization events may be a biological feature that extends across other lineages within the *C. neoformans* and *C. gattii* species complexes (Farrer *et al.* 2015; Hagen *et al.* 2015), prompting a need for wider investigation of the population genomic structure of the entire complex using a rigorously applied GCPSR framework to support formal changes in taxonomy (Kwon-Chung *et al.* 2017).

The placement of isolates from Brazil at basal branching positions of the two VNB subclades phylogenetically separates the South American and African isolates within both the VNBI and VNBII groups. This finding, along with the presence of a large number of unique alleles in each of these four subclades and strong haplotype sharing seen with fineSTRUCTURE analysis (Figure 6 and Figure 7), suggests that there were ancient migrations of the VNB group between Africa and South America following the initial divergence of VNBI and VNBII, but prior to each group's radiation. This finding appears consistent with a prior report of diverse isolates from Brazil in a new VNI genotype 1B (de Oliveira *et al.* 2004). While the lack of a trustworthy molecular clock combined with substantial rates of recombination currently precludes confidently dating the time of divergence between VNB from South America and Africa, this division clearly occurred after

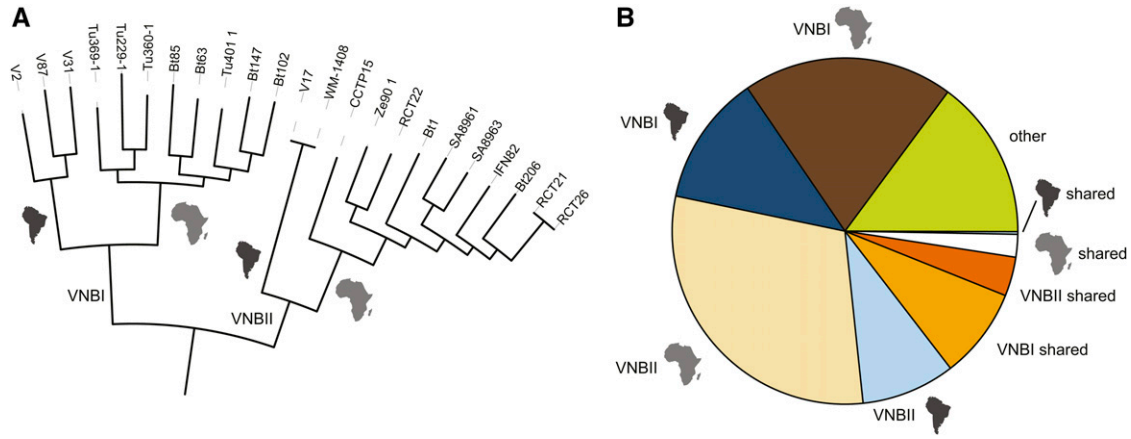

Figure 6 VNB alleles in population subdivisions and across geography. (A) Phylogeny of VNB lineage showing major subdivisions (VNBI and VNBII) and inferred ancestral geography (South America or Africa, depicted as continent shapes). (B) Classification of all 445,193 private VNB alleles (present in at least one VNB isolate and no VNI or VNII isolates) by subdivisions and geography. Most VNB alleles are specific for each VNB subdivision and for the geographic subdivisions within each group. More alleles are shared between geographic locations in the same subdivision (VNBI or VNBII) than are shared within geographic locations across subdivisions.

these continents split over 110 MYA, and also after VNB itself subdivided into two lineages—VNBI and VNBII. As is the case with VNI, cross-Atlantic migration events may also have vectored VNB between these two continents. Despite evidence for these migration events, the majority of VNI and VNII migrations were likely much more recent than is seen with VNB, with nearly clonal isolates of VNI and VNII found in disparate geographic regions. The presence of one South American VNB isolate (V53) that nests within African isolates on the phylogeny suggests a limited number of more recent migration events may be occurring between the two regions even within VNB, despite the large degree of reproductive isolation that we observed. Identification of additional South American VNB isolates is necessary to determine their diversity and relationship to isolates from African continental regions. Although the sequenced isolates all contain the *MAT*α genotype, our sample size was small and likely underrepresents the true diversity of this lineage in South America and the ecological reservoirs that it occupies.

Given the propensity of *C. neoformans* var. *grubii* VNI and VNII for having an environmental reservoir in bird excreta [unlike VNB which is principally associated with arboreal reservoirs (Litvintseva *et al.* 2011; Vanhove *et al.* 2017)], it has been proposed that radiations of birds, likely pigeons, globally dispersed *C. neoformans* var. *grubii* from a genetically diverse population in southern Africa (Litvintseva *et al.* 2011), resulting in an expansion of the *C. neoformans* var. *grubii* VNI out of Africa. Litvintseva *et al.* (2011) hypothesized that this "out-of-Africa" model for the evolution of VNI explains the origin of the global VNI population. Other studies showing lower genetic diversity of VNI populations in Southeast Asia (Simwami *et al.* 2011) and in South America (Ferreira-Paim *et al.* 2017) further support an African origin of *C. neoformans* var. *grubii*. An alternative explanation for the higher diversity of African VNI could be that this lineage originated elsewhere and became more diverse in this continent by mating with the "native" VNB population or due to other factors. Our analysis did not find a large subset of VNB alleles within the African VNI isolates based on ancestry analysis. In addition, we found one VNI subclade comprised mostly of African isolates that appears to be recombining at a higher frequency than other VNI groups. The phylogenetic intermixing of isolates from India and Africa strongly support the hypothesis that there is long-range dispersal and ancient recombination in environmental populations in India and Africa, indicative of multiple migratory events across time and into the present. Did VNI therefore evolve out of Africa? Further sampling of environmental isolates from across South America and more diverse regions of Africa, as well as correct estimation of the mutation rate in *C. neoformans* var. *grubii* to allow calibration of a molecular clock, is needed to further test this hypothesis.

While gene content is very similar across the *C. neoformans* var. *grubii* lineages, we found examples of lineage-specific genes including clusters unique to VNI or VNII. While this suggests that the *C. neoformans* var. *grubii* gene inventory based on H99 (Janbon *et al.* 2014) is largely representative of all lineages, additional genes specific to VNII and VNB are important to consider in studies focusing on isolates of these lineages. Differences in gene expression may also differentiate the lineages, and it is important to note that these will include lineage-specific genes that may contribute to variation in clinical profiles and virulence that occur among lineages of *C. neoformans* var. *grubii* (Beale *et al.* 2015). In addition, we found the most rapidly evolving genes between each of the lineages include transcription factors and transferases, suggesting phenotypic diversity may be generated through transcriptional reprogramming and protein modification rather than changes in gene content. The *SXI1* gene detected in comparisons of VNII with both VNI and VNB appears to be highly substituted in the VNII lineage; this sequence divergence of *SXI1* in VNII could contribute to

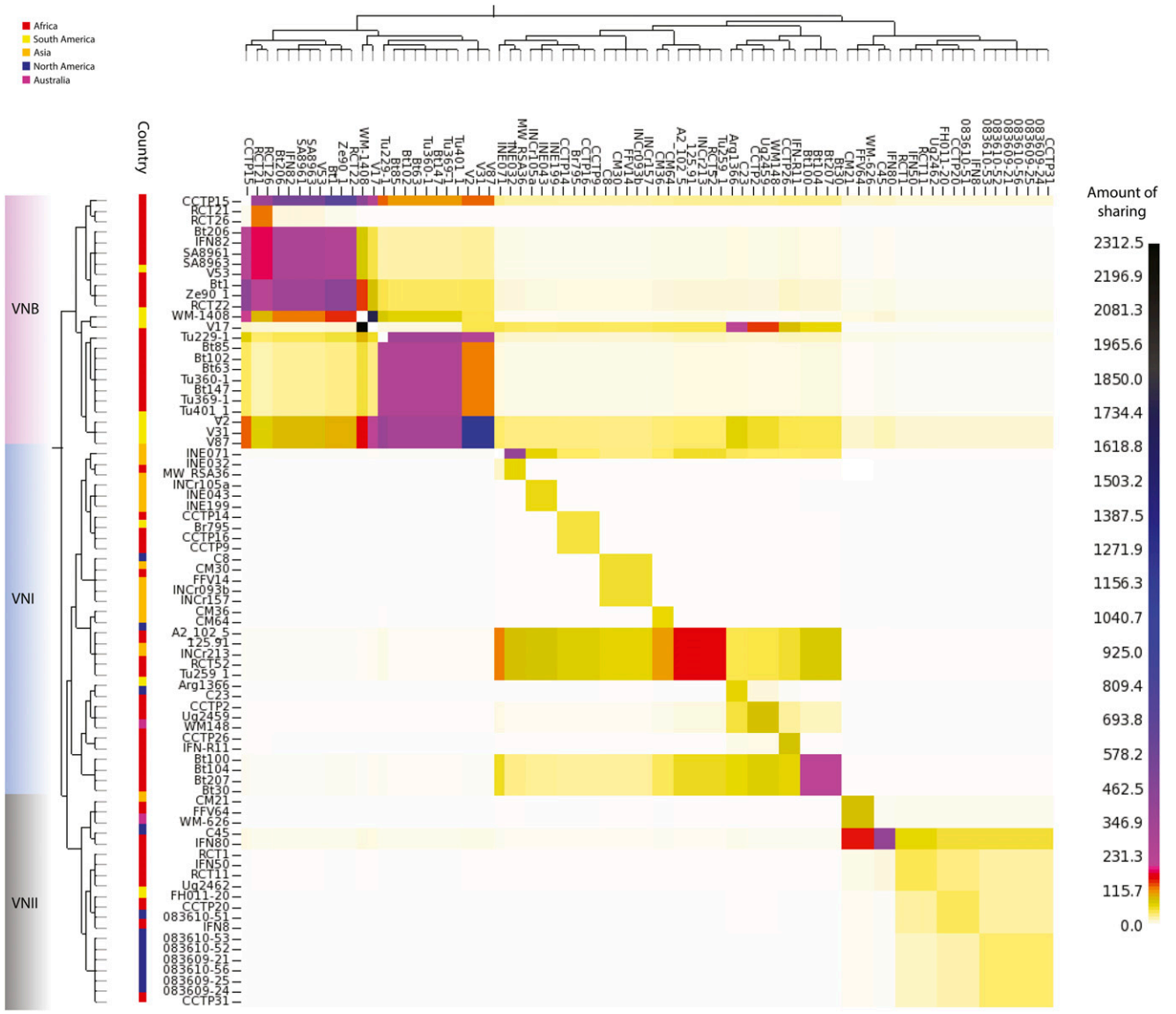

Figure 7 Genome-sharing analysis of *C. neoformans* var. *grubii* using fineSTRUCTURE was performed on a SNP matrix using a representative of each clonal population within the VNI lineage. These genomes were reduced to a pairwise similarity matrix, which facilitates the identification of population structure based on haplotype sharing within regions of the genome. The *x*-axis represents the "donor" of genomic regions, while the *y*-axis represents the recipient of shared genomic regions. The scale bar represents the amount of genomic sharing, with black representing the largest amount of sharing of genetic material, and white representing the least amount of shared genetic material (no sharing). The geographic site of isolation is illustrated with colored boxes as in Figure 1, and lineages are also shown.

differences in mating with this group. Truncated alleles of *SXI1* are frequently observed in the serotype-D *MAT*α chromosome of AD hybrids and are suggested to contribute to increased mating efficiency (Lin *et al.* 2007).

Our analysis revealed that hybrid isolates originate from multiple lineages and resolved the parental genotypes. Prior analysis with MLST loci suggested that some isolates contain a mix of multiple genotypes (Litvintseva *et al.* 2003; Chen *et al.* 2015). However, the sensitivity and precision of these methods has been limited by the small number of loci examined, the use of genes involved in virulence that may be under different selective pressure, as well as incomplete lineage sorting in some groups. Analysis of genome-wide variation revealed that some isolates appear to be a recent mix of different ancestries, based on the detection of large blocks of sites with each ancestry; this could result from a small number of crossing over events for each chromosome during meiosis. Other isolates contain more highly intermixed ancestry across the genome and are predominantly of a single ancestry; these may have occurred by more historical hybridization followed by subsequent mating within a single lineage group. The demonstration of genome mixing in hybrid isolates raises interesting questions about whether such fundamentally new assortments of the three lineages could generate genotypes with new phenotypes, which perhaps have a fitness or selective advantage.

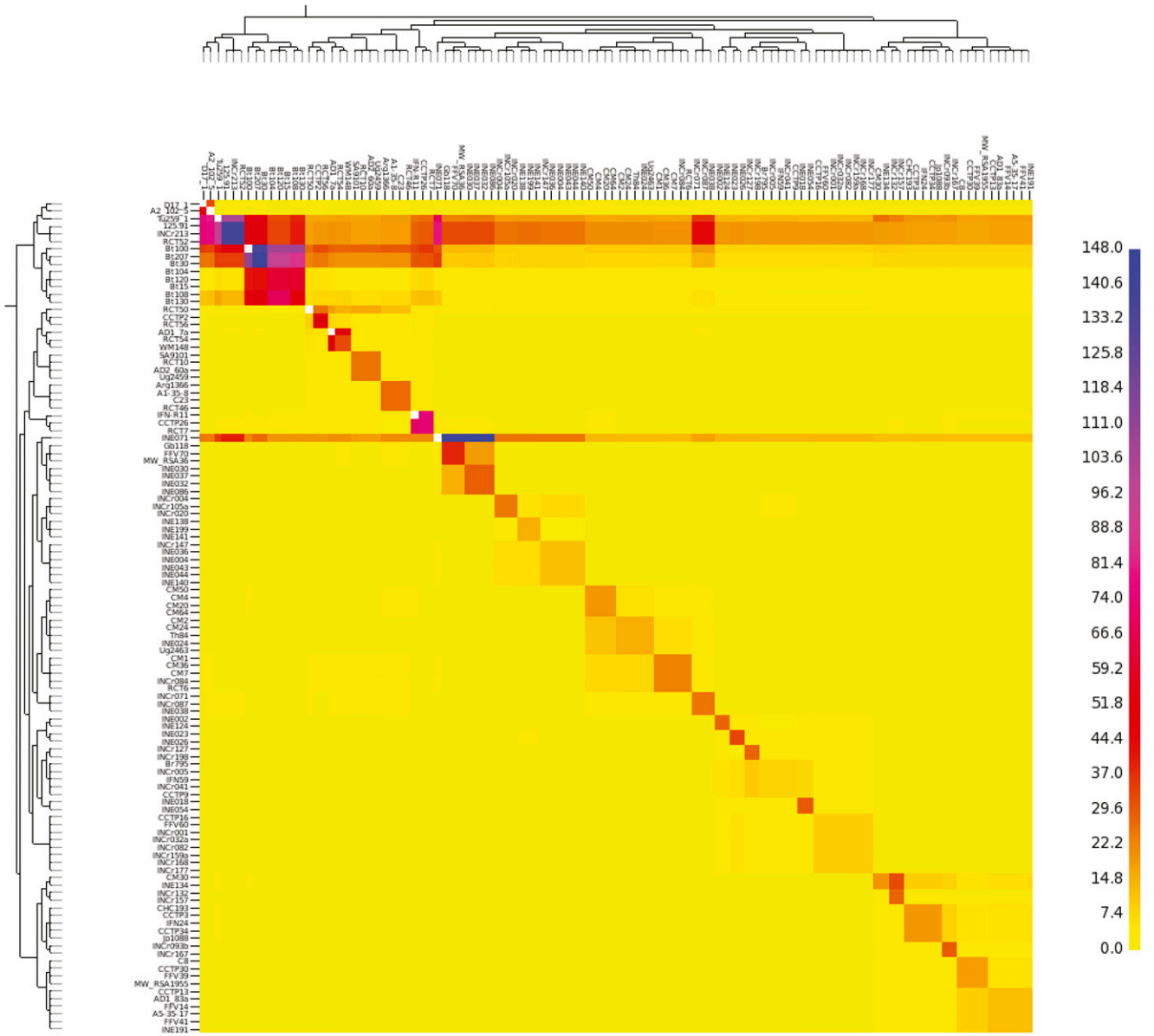

Figure 8 Genome-sharing analysis of the VNI lineage using fineSTRUCTURE on a SNP matrix of 111 genomes. The *x*-axis represents the donor of genomic regions, while the *y*-axis represents the recipient of shared genomic regions. The scale bar represents the amount of genomic sharing, with blue representing the largest amount of sharing of genetic material, and yellow representing the least amount of shared genetic material (no sharing).

Analysis of hybrids between serotypes A and D revealed a remarkable degree of genome reassortment. All of the eight sequenced AD isolates show evidence of aneuploidy, affecting the copy number of 12 of 14 serotype A-derived chromosomes and all 14 serotype D-derived chromosomes. This is consistent with the high rate of AD-isolate aneuploidy previously reported using flow analysis of DNA content (Lengeler *et al.* 2001) or comparative genome hybridization (Hu *et al.* 2008). For some chromosomes, only one parental genotype was detected in a subset of five isolates; this includes a loss of the serotype-D copy of chromosome 1, as previously observed in analysis of three AD hybrid isolates (Hu *et al.* 2008). However, we further find that LOH in some cases is due to partial copies of several chromosomes, suggesting that genomic instability in AD hybrids may result in chromosomal breakage. LOH was also observed for smaller regions in diploid AA hybrids. Similar LOH events are frequently observed in diploid fungi including *C. albicans* (Hirakawa *et al.* 2015) and may contribute to the generation of genetic diversity in both species.

Aneuploidy was also commonly observed in the haploid *C. neoformans* var. *grubii* isolates. Additional copies of regions of chromosome 1 which include *AFR1* or *ERG11* are associated with drug resistance, though aneuploidies of additional chromosomes are also observed (Sionov *et al.* 2010). Although functional significance of aneuploidy of other chromosomes is less well understood, most of the isolates exhibiting aneuploidy are of clinical origin, suggesting an

increased copy of other genes may provide an advantage or that there is higher genome instability during infection. An isochromosome of the left arm of chromosome 12 that arose during infection has been reported (Ormerod *et al.* 2013), and chromosome 12 aneuploidy is widely seen in African patients with relapsed infections (Chen *et al.* 2017; Rhodes *et al.* 2017), although the specific role of this duplication is unclear. Our data suggest that there could be additional isochromosomes based on the detection of partial chromosomes using sequencing read depth. Alternatively, these regions could be represented in the genome as fusions with other chromosomes.

Previous studies of *C. gattii* have pointed toward South America as a source of the diversity for the *C. gattii* VGII lineage (Hagen *et al.* 2013; Engelthaler *et al.* 2014). Given the shared evolutionary history of *C. gattii* and *C. neoformans* var. *grubii* (Xu *et al.* 2000), South America could also represent a major diversity center of *C. neoformans* var. *grubii*. Our data suggest that *C. neoformans* var. *grubii* VNB isolates in both subgroups from South America have undergone ancestral recombination events, donating genetic material to all lineages across multiple geographic locations. Our study also provides clear evidence that recombination is more limited by lineage than by geographic barriers; the transcontinental nature of *C. neoformans* var. *grubii*, particularly the VNI and VNII lineages, supports the hypothesis of historical or ongoing migration events to facilitate such recombination. Our study identified recombination within the VNI and VNII lineages, where nearly all the isolates contain the *MAT*α mating type. This suggests that mating likely occurs between *MAT*α isolates, as is found in *C. neoformans* var. *neoformans* (Sun *et al.* 2014). Previous studies have hypothesized that *C. neoformans* var. *grubii* can complete its sexual reproductive life cycle in environmental niches, such as plants (Xue *et al.* 2007) and pigeon guano (Nielsen *et al.* 2007; Vanhove *et al.* 2017). Our observations that all lineages of *C. neoformans* var. *grubii* show the ability to widely disperse, recombine, and hybridize, across large geographic distances, illustrates that this pathogen has a high degree of evolutionary plasticity. Therefore, lineages that have not drifted in the frequency of their mating types are likely to display higher rates of recombination and hybridization. These factors are likely related to the success of *C. neoformans* var. *grubii* in infecting the immunosuppressed "human environment," thereby causing a high burden of mortality worldwide (Armstrong-James *et al.* 2014).

Acknowledgments

We thank Jose Muñoz for helpful comments on the manuscript and the Broad Institute Genomics Platform for generating DNA sequences for this study. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant number U19 AI-110818 and by the National Human Genome Research Institute grant number U54HG003067 to the Broad Institute. Support to J.R.P. came from Public Health Service grants AI73896 and AI93257. J.R. and M.A.B. were supported by a United Kingdom Medical Research Council (MRC) grant awarded to M.C.F., T.B., and T.S.H. (MRC MR/K000373/1). M.V. was supported by a United Kingdom Natural Environment Research Council Ph.D. studentship. J.H. was supported by National Institutes of Health grants AI-39115-19 and AI-50113-13. D.J.L. was funded by Wellcome Trust grant number WT104125MA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: Investigation: J.R., C.A.D., S.M.S., S. Sakthikumar, and C.A.C. Validation: J.R., C.A.D., S.M.S., and C.A.C. Visualization: J.R., C.A.D., S.M.S., S. Sakthikumar, and C.A.C. Writing, original draft preparation: J.R., C.A.D., and C.A.C. Writing, review and editing: J.R., C.A.D., M.C.F., C.A.C., A.A., M.A.B., D.M.E., W.M., F.H., J.-M.V., J.H., A.L., and J.R.P. Resources: M.C.F., C.A.C., S. Saif, S.G., M.V., Y.C., J.R.P., T.B., T.S.H., V.P., A.L.C., A.C., F.H., M.T.I.-Z., W.M., D.M.E., A.A., J.-M.V., J.H., and D.J.L. Supervision: C.A.C. and M.C.F. Funding acquisition: C.A.C. and M.C.F. Conceptualization: C.A.C., M.C.F., and A.L.C.

Literature Cited

Armstrong-James, D., G. Meintjes, and G. D. Brown, 2014 A neglected epidemic: fungal infections in HIV/AIDS. Trends Microbiol. 22: 120–127.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. J. Comput. Mol. Cell Biol. 19: 455–477.

Beale, M. A., W. Sabiiti, E. J. Robertson, K. M. Fuentes-Cabrejo, S. J. O'Hanlon *et al.*, 2015 Genotypic diversity is associated with clinical outcome and phenotype in Cryptococcal meningitis across Southern Africa. PLoS Negl. Trop. Dis. 9: e0003847.

Bicanic, T., T. Harrison, A. Niepieklo, N. Dyakopu, and G. Meintjes, 2006 Symptomatic relapse of HIV-associated Cryptococcal meningitis after initial fluconazole monotherapy: the role of fluconazole resistance and immune reconstitution. Clin. Infect. Dis. 43: 1069–1070.

Bicanic, T., G. Meintjes, R. Wood, M. Hayes, K. Rebe *et al.*, 2007 Fungal burden, early fungicidal activity, and outcome in cryptococcal meningitis in antiretroviral-naive or antiretroviral-experienced patients treated with amphotericin B or fluconazole. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 45: 76–80.

Bicanic, T., R. Wood, G. Meintjes, K. Rebe, A. Brouwer *et al.*, 2008 High-dose amphotericin B with flucytosine for the treatment of cryptococcal meningitis in HIV-infected patients: a randomized trial. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 47: 123–130.

Bovers, M., F. Hagen, E. E. Kuramae, and T. Boekhout, 2008 Six monophyletic lineages identified within *Cryptococcus neoformans* and *Cryptococcus gattii* by multi-locus sequence typing. Fungal Genet. Biol. FG B 45: 400–421.

Brouwer, A. E., A. Rajanuwong, W. Chierakul, G. E. Griffin, R. A. Larsen *et al.*, 2004 Combination antifungal therapies for

HIV-associated cryptococcal meningitis: a randomised trial. Lancet Lond. Engl. 363: 1764–1767.

Bui, T., X. Lin, R. Malik, J. Heitman, and D. Carter, 2008 Isolates of *Cryptococcus neoformans* from infected animals reveal genetic exchange in unisexual, alpha mating type populations. Eukaryot. Cell 7: 1771–1780.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón, 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973.

Casadevall, A., and J. R. Perfect, 1998 *Cryptococcus neoformans*. ASM Press, Washington, DC.

Chen, Y., A. P. Litvintseva, A. E. Frazzitta, M. R. Haverkamp, L. Wang *et al.*, 2015 Comparative analyses of clinical and environmental populations of *Cryptococcus neoformans* in Botswana. Mol. Ecol. 24: 3559–3571.

Chen, Y., R. A. Farrer, C. Giamberardino, S. Sakthikumar, A. Jones *et al.*, 2017 Microevolution of serial clinical isolates of *Cryptococcus neoformans* var. *grubii* and *C. gattii*. MBio 8: e00166–e17.

Cogliati, M., 2013 Global molecular epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: an atlas of the molecular types. Scientifica 2013: 675213.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Day, J. N., 2004 Cryptococcal meningitis. Pract. Neurol. 4: 274–285.

de Oliveira, M. T. B., T. Boekhout, B. Theelen, F. Hagen, F. A. Baroni *et al.*, 2004 *Cryptococcus neoformans* shows a remarkable genotypic diversity in Brazil. J. Clin. Microbiol. 42: 1356–1359.

Desjardins, C. A., C. Giamberardino, S. M. Sykes, C.-H. Yu, J. L. Tenor *et al.*, 2017 Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 27: 1207–1219.

Desnos-Ollivier, M., S. Patel, D. Raoux-Barbot, J. Heitman, F. Dromer *et al.*, 2015 Cryptococcosis serotypes impact outcome and provide evidence of *Cryptococcus neoformans* speciation. MBio 6: e00311.

Dettman, J. R., D. J. Jacobson, and J. W. Taylor, 2003 A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. Evolution 57: 2703–2720.

Dromer, F., S. Mathoulin-Pélissier, O. Launay, and O. Lortholary French Cryptococcosis Study Group, 2007 Determinants of disease presentation and outcome during cryptococcosis: the CryptoA/D study. PLoS Med. 4: e21.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Engelthaler, D. M., N. D. Hicks, J. D. Gillece, C. C. Roe, J. M. Schupp *et al.*, 2014 *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. MBio 5: e01464-14.

Esher, S. K., K. S. Ost, L. Kozubowski, D. H. Yang, M. S. Kim *et al.*, 2016 Relative contributions of prenylation and postprenylation processing in *Cryptococcus neoformans* pathogenesis. mSphere 1: e00084-15.

Farrer, R. A., C. A. Desjardins, S. Sakthikumar, S. Gujja, S. Saif *et al.*, 2015 Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. MBio 6: e00868-15.

Ferreira-Paim, K., L. Andrade-Silva, F. M. Fonseca, T. B. Ferreira, D. J. Mora *et al.*, 2017 MLST-based population genetic analysis in a global context reveals clonality amongst *Cryptococcus neoformans* var. *grubii* VNI isolates from HIV patients in Southeastern Brazil. PLoS Negl. Trop. Dis. 11: e0005223.

Franzot, S. P., I. F. Salkin, and A. Casadevall, 1999 *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. J. Clin. Microbiol. 37: 838–840.

Gerstein, A. C., M. S. Fu, L. Mukaremera, Z. Li, K. L. Ormerod *et al.*, 2015 Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. MBio 6: e01340-15.

Gilbert, N. M., M. J. Donlin, K. J. Gerik, C. A. Specht, J. T. Djordjevic *et al.*, 2010 KRE genes are required for β-1,6-glucan synthesis, maintenance of capsule architecture and cell wall protein anchoring in *Cryptococcus neoformans*. Mol. Microbiol. 76: 517–534.

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9: R7.

Hagen, F., P. C. Ceresini, I. Polacheck, H. Ma, F. van Nieuwerburgh *et al.*, 2013 Ancient dispersal of the human fungal pathogen *Cryptococcus gattii* from the Amazon rainforest. PLoS One 8: e71148.

Hagen, F., K. Khayhan, B. Theelen, A. Kolecka, I. Polacheck *et al.*, 2015 Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet. Biol. 78: 16–48.

Heitman, J., T. R. Kozel, K. J. Kwon-Chung, J. R. Perfect, and A. Casadevall, 2011 *Cryptococcus: From Human Pathogen to Model Yeast*. ASM Press, Washington, DC.

Hirakawa, M. P., D. A. Martinez, S. Sakthikumar, M. Z. Anderson, A. Berlin *et al.*, 2015 Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res. 25: 413–425.

Hiremath, S. S., A. Chowdhary, T. Kowshik, H. S. Randhawa, S. Sun *et al.*, 2008 Long-distance dispersal and recombination in environmental populations of *Cryptococcus neoformans* var. *grubii* from India. Microbiology 154: 1513–1524.

Hu, G., I. Liu, A. Sham, J. E. Stajich, F. S. Dietrich *et al.*, 2008 Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. Genome Biol. 9: R41.

Hu, G., J. Wang, J. Choi, W. H. Jung, I. Liu *et al.*, 2011 Variation in chromosome copy number influences the virulence of *Cryptococcus neoformans* and occurs in isolates from AIDS patients. BMC Genomics 12: 526.

Hull, C. M., R. C. Davidson, and J. Heitman, 2002 Cell identity and sexual development in *Cryptococcus neoformans* are controlled by the mating-type-specific homeodomain protein Sxi1alpha. Genes Dev. 16: 3046–3060.

Inglis, D. O., M. S. Skrzypek, E. Liaw, V. Moktali, G. Sherlock *et al.*, 2014 Literature-based gene curation and proposed genetic nomenclature for *Cryptococcus*. Eukaryot. Cell 13: 878–883.

Janbon, G., K. L. Ormerod, D. Paulet, E. J. Byrnes, V. Yadav *et al.*, 2014 Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. 10: e1004261.

Jarvis, J. N., G. Meintjes, K. Rebe, G. N. Williams, T. Bicanic *et al.*, 2012 Adjunctive interferon-γ immunotherapy for the treatment of HIV-associated cryptococcal meningitis: a randomized controlled trial. AIDS 26: 1105–1113.

Jung, K.-W., D.-H. Yang, S. Maeng, K.-T. Lee, Y.-S. So *et al.*, 2015 Systematic functional profiling of transcription factor networks in *Cryptococcus neoformans*. Nat. Commun. 6: 6757.

Kavanaugh, L. A., J. A. Fraser, and F. S. Dietrich, 2006 Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-Gene fragment. Mol. Biol. Evol. 23: 1879–1890.

Khayhan, K., F. Hagen, W. Pan, S. Simwami, M. C. Fisher *et al.*, 2013 Geographically structured populations of *Cryptococcus*

neoformans variety *grubii* in Asia correlate with HIV status and show a clonal population structure. PLoS One 8: e72222.

Kwon-Chung, K. J., 1975 A new genus, *filobasidiella*, the perfect state of *Cryptococcus neoformans*. Mycologia 67: 1197–1200.

Kwon-Chung, K. J., 1976 A new species of *Filobasidiella*, the sexual state of *Cryptococcus neoformans* B and C serotypes. Mycologia 68: 943–946.

Kwon-Chung, K. J., J. E. Bennett, B. L. Wickes, W. Meyer, C. A. Cuomo *et al.* 2017 The case for adopting the "species complex" nomenclature for the etiologic agents of Cryptococcosis. mSphere 2: e00357-e16.

Lawson, D. J., G. Hellenthal, S. Myers, and D. Falush, 2012 Inference of population structure using dense haplotype data. PLoS Genet. 8: e1002453.

Lee, D.-J., Y.-S. Bahn, H.-J. Kim, S.-Y. Chung, and H. A. Kang, 2015 Unraveling the novel structure and biosynthetic pathway of *O*-linked glycans in the Golgi apparatus of the human pathogenic yeast *Cryptococcus neoformans*. J. Biol. Chem. 290: 1861–1873.

Lengeler, K. B., R. C. Davidson, C. D'Souza, T. Harashima, W. C. Shen *et al.*, 2000a Signal transduction cascades regulating fungal development and virulence. Microbiol. Mol. Biol. Rev. 64: 746–785.

Lengeler, K. B., P. Wang, G. M. Cox, J. R. Perfect, and J. Heitman, 2000b Identification of the MATa mating-type locus of *Cryptococcus neoformans* reveals a serotype A MATa strain thought to have been extinct. Proc. Natl. Acad. Sci. USA 97: 14455–14460.

Lengeler, K. B., G. M. Cox, and J. Heitman, 2001 Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect. Immun. 69: 115–122.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. Available at: https://arxiv.org/abs/1303.3997.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Li, L., C. J. Stoeckert, and D. S. Roos, 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13: 2178–2189.

Lin, X., C. M. Hull, and J. Heitman, 2005 Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434: 1017–1021.

Lin, X., A. P. Litvintseva, K. Nielsen, S. Patel, A. Floyd *et al.*, 2007 αADα hybrids of *Cryptococcus neoformans*: evidence of same-sex mating in nature and hybrid fitness. PLoS Genet. 3: e186.

Lin, X., S. Patel, A. P. Litvintseva, A. Floyd, T. G. Mitchell *et al.*, 2009 Diploids in the *Cryptococcus neoformans* serotype A population homozygous for the α mating type originate via unisexual mating. PLoS Pathog. 5: e1000283.

Litvintseva, A. P., and T. G. Mitchell, 2012 Population genetic analyses reveal the African Origin and strain variation of *Cryptococcus neoformans* var. *grubii*. PLoS Pathog. 8: e1002495.

Litvintseva, A. P., R. E. Marra, K. Nielsen, J. Heitman, R. Vilgalys *et al.*, 2003 Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. Eukaryot. Cell 2: 1162–1168.

Litvintseva, A. P., L. Kestenbaum, R. Vilgalys, and T. G. Mitchell, 2005 Comparative analysis of environmental and clinical populations of *Cryptococcus neoformans*. J. Clin. Microbiol. 43: 556–564.

Litvintseva, A. P., R. Thakur, R. Vilgalys, and T. G. Mitchell, 2006 Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. Genetics 172: 2223–2238.

Litvintseva, A. P., X. Lin, I. Templeton, J. Heitman, and T. G. Mitchell, 2007 Many globally isolated AD hybrid strains of *Cryptococcus neoformans* originated in Africa. PLoS Pathog. 3: e114.

Litvintseva, A. P., I. Carbone, J. Rossouw, R. Thakur, N. P. Govender *et al.*, 2011 Evidence that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have evolved in Africa. PLoS One 6: e19688.

Liu, O. W., C. D. Chun, E. D. Chow, C. Chen, H. D. Madhani *et al.*, 2008 Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. Cell 135: 174–188.

Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo *et al.*, 2005 The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science 307: 1321–1324.

Loyse, A., D. Wilson, G. Meintjes, J. N. Jarvis, T. Bicanic *et al.*, 2012 Comparison of the early fungicidal activity of high-dose fluconazole, voriconazole, and flucytosine as second-line drugs given in combination with amphotericin B for the treatment of HIV-associated cryptococcal meningitis. Clin. Infect. Dis. 54: 121–128.

Lugarini, C., C. S. Goebel, L. A. Z. Condas, M. D. Muro, M. R. de Farias *et al.*, 2008 *Cryptococcus neoformans* isolated from Passerine and Psittacine bird excreta in the state of Paraná, Brazil. Mycopathologia 166: 61–69.

Maccallum, I., D. Przybylski, S. Gnerre, J. Burton, I. Shlyakhter *et al.*, 2009 ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol. 10: R103.

Maziarz, E. K., and J. R. Perfect, 2016 Cryptococcosis. Infect. Dis. Clin. North Am. 30: 179–206.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meyer, W., K. Marszewska, M. Amirmostofian, R. P. Igreja, C. Hardtke *et al.*, 1999 Molecular typing of global isolates of *Cryptococcus neoformans* var. *neoformans* by polymerase chain reaction fingerprinting and randomly amplified polymorphic DNA-a pilot study to standardize techniques on which to base a detailed epidemiological survey. Electrophoresis 20: 1790–1799.

Meyer, W., D. M. Aanensen, T. Boekhout, M. Cogliati, M. R. Diaz *et al.*, 2009 Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med. Mycol. 47: 561–570.

Ngamskulrungroj, P., F. Gilgado, J. Faganello, A. P. Litvintseva, A. L. Leal *et al.*, 2009 Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS One 4: e5862.

Ni, M., M. Feretzaki, W. Li, A. Floyd-Averette, P. Mieczkowski *et al.*, 2013 Unisexual and heterosexual meiotic reproduction generate aneuploidy and phenotypic diversity de novo in the yeast *Cryptococcus neoformans*. PLoS Biol. 11: e1001653.

Nielsen, K., A. L. D. Obaldia, and J. Heitman, 2007 *Cryptococcus neoformans* mates on pigeon guano: implications for the realized ecological niche and globalization. Eukaryot. Cell 6: 949–959.

Ormerod, K. L., C. A. Morrow, E. W. L. Chow, I. R. Lee, S. D. M. Arras *et al.*, 2013 Comparative genomics of serial isolates of *Cryptococcus neoformans* reveals gene associated with carbon utilization and virulence. G3 3: 675–686.

Park, B. J., K. A. Wannemuehler, B. J. Marston, N. Govender, P. G. Pappas *et al.*, 2009 Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. AIDS 23: 525–530.

Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. PLoS Genet. 2: e190.

Pfeifer, B., U. Wittelsbürger, S. E. Ramos-Onsins, and M. J. Lercher, 2014 PopGenome: an efficient swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31: 1929–1936.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Rajasingham, R., S. Smith, B. Park, J. Jarvis, N. Govender *et al.*, 2017 Global burden of disease of HIV-associated Cryptococcal meningitis: an updated analysis. Lancet Infect. Dis. DOI: 10.1016/S1473-3099(17)30243-8.

Rhodes, J., M. A. Beale, M. Vanhove, J. N. Jarvis, S. Kannambath *et al.*, 2017 A population genomics approach to assessing the genetic basis of within-host microevolution underlying recurrent Cryptococcal meningitis infection. G3 7: 1165–1176.

Selvig, K., E. R. Ballou, C. B. Nichols, and J. A. Alspaugh, 2013 Restricted substrate specificity for the geranylgeranyltransferase-I enzyme in *Cryptococcus neoformans*: implications for virulence. Eukaryot. Cell 12: 1462–1471.

Simwami, S. P., K. Khayhan, D. A. Henk, D. M. Aanensen, T. Boekhout *et al.*, 2011 Low diversity *Cryptococcus neoformans* variety *grubii* multilocus sequence types from Thailand are consistent with an ancestral African origin. PLoS Pathog. 7: e1001343.

Sionov, E., H. Lee, Y. C. Chang, and K. J. Kwon-Chung, 2010 *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. PLoS Pathog. 6: e1000848.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Sun, S., R. B. Billmyre, P. A. Mieczkowski, and J. Heitman, 2014 Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. PLoS Genet. 10: e1004849.

Taylor, J. W., D. J. Jacobson, S. Kroken, T. Kasuga, D. M. Geiser *et al.*, 2000 Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31: 21–32.

Ter-Hovhannisyan, V., A. Lomsadze, Y. O. Chernoff, and M. Borodovsky, 2008 Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18: 1979–1990.

Vanhove, M., M. A. Beale, J. Rhodes, D. Chanda, S. Lakhi *et al.*, 2017 Genomic epidemiology of *Cryptococcus* yeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. Mol. Ecol. 26: 1991–2005.

Viviani, M. A., M. C. Esposto, M. Cogliati, M. T. Montagna, and B. L. Wickes, 2001 Isolation of a *Cryptococcus neoformans* serotype A MATa strain from the Italian environment. Med. Mycol. 39: 383–386.

Vogan, A. A., and J. Xu, 2014 Evidence for genetic incompatibilities associated with post-zygotic reproductive isolation in the human fungal pathogen *Cryptococcus neoformans*. Genome 57: 335–344.

Wang, L., B. Zhai, and X. Lin, 2012 The link between morphotype transition and virulence in *Cryptococcus neoformans*. PLoS Pathog. 8: e1002765.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358–1370.

Xu, J., R. Vilgalys, and T. G. Mitchell, 2000 Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. Mol. Ecol. 9: 1471–1481.

Xue, C., Y. Tada, X. Dong, and J. Heitman, 2007 The human fungal pathogen *Cryptococcus* can complete its sexual cycle during a pathogenic association with plants. Cell Host Microbe 1: 263–273.

Yahara, K., Y. Furuta, K. Oshima, M. Yoshida, T. Azuma *et al.*, 2013 Chromosome painting in silico in a bacterial species reveals fine population structure. Mol. Biol. Evol. 30: 1454–1464.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591.

Communicating editor: A. P. Mitchell